# Work with What You've Got: Improving Teachers' Pedagogical Skills at Scale in Rural Peru[1]


| Juan F. Castro | Paul Glewwe | Alexandra Heredia-Mayo | Ricardo Montero |
|---|---|---|---|
| Universidad del Pacifico | University of Minnesota | Universidad del Pacifico | University of Minnesota |


This version: September 2021


**Abstract**

We evaluate the impact of a large-scale teacher coaching program, in a context of high teacher turnover, on teachers' pedagogical skills. Previous studies find that small-scale coaching programs can improve teaching of reading and science in developing countries. However, scaling up can reduce programs' effectiveness, and teacher turnover can erode compliance and cause spillovers onto non-program schools. We develop a framework that defines different treatment effects when teacher turnover is present, and explains which effects can be estimated. We evaluate a Peruvian teacher coaching program, exploiting random assignment of that program's expansion to 3,795 rural schools in 2016. After two years, teachers assigned to the program increased their aggregate pedagogical skills by 0.20 standard deviations. The program also increased student learning, and the schools whose teachers' pedagogical skills increase the most are also the schools with the highest increases in learning, indicating that pedagogical skills are one mechanism linking teacher coaching to student learning.



Keywords: teacher coaching, pedagogical skill, teacher turnover.
JEL Codes: I21, O15.



[1] We would like to thank seminar participants at the Department of Applied Economics of the University of Minnesota, the Department of Economics of Universidad del Rosario, the LACEA 2019 Annual Meeting, the Department of Agricultural and Consumer Economics at the University of Illinois, and the Department of Agricultural Economics and Rural Development at Seoul National University for their valuable comments. We are also grateful to Hugo Fernández for excellent research assistance. Any remaining errors are ours alone.

# 1. Introduction

Teacher quality is an essential determinant of student learning (Das et al. 2007, Clotfelter et al. 2010, Chetty et al. 2014). Yet many teachers lack mastery in the subjects they teach, or lack the pedagogical skills to teach them effectively. This is especially true for teachers in developing countries (World Bank, 2018). Can these teachers' skills be improved?

Every year, developing countries spend over $1 billion on teacher training (Loyalka et al., 2019). Popova et al. (2016) find that about two thirds of the World Bank educational projects between 2000 and 2012 included in-service teacher training. Such training is attractive because it can be centrally designed and coordinated by the Ministry of Education and is usually supported by teachers' unions (Evans and Popova, 2016).

In this study, we evaluate the impact of a large-scale teacher coaching program, operating in a context of high teacher turnover, on a broad range of pedagogical skills. Evidence on the impacts of in-service training in developing countries is mixed, and programs vary widely in form and content. A survey by Evans and Popova (2016) found that programs with face-to-face training, follow-up visits, engagement of teachers to obtain their ideas, and adaption to local context, tend to have larger effects on student learning. Coaching programs often have these features as they involve school visits, classroom observations, and personalized feedback for teachers by trained peers or coaches. Thus, coaching programs are a promising alternative to traditional in-service training that offers intensive sessions to large numbers of teachers at a centralized venue.

When programs are offered at the school level but are intended to operate through teachers, and teachers can move between schools, estimates of the effectiveness of the program based on a randomized control trial may be biased. In particular, movement of teachers across schools may lead to spillovers that will introduce biases when comparing treated and control schools, even when all schools comply with their random assignment and there are no biases due to the selection or attrition of students.

Education interventions that operate through teachers often have all teachers in a school share treatment status (i.e., all teachers are either treated or untreated), Most studies of the effectiveness of these types of interventions focus on student outcomes and compare treatment with control schools, and some of them evaluate results after enough time has passed for teachers to switch schools (Lucas et. al. 2014, Jukes et. al. 2017, Cilliers et. al. 2020). These studies usually address potential biases due to student attrition, yet they rarely mention the possibility of teacher turnover or the potential bias it may induce.

This risk of bias may occur not only for education interventions but also for any estimation of treatment effects in cluster randomized control trials (RCTs) with movement of service providers or program beneficiaries across clusters. Indeed, high turnover is reported for many non-education contexts. For example, Kovner et al. (2014) report that 17.5% of new nurses in the U.S. leave their jobs within one year of starting, and Banerjee et al. (2021) find, in their control sample, that one-third of police officers in India changed stations over an 18-month period. Despite being quite frequent, turnover is usually ignored in program evaluations. For example, Georgiadis and Pitellis (2016) compare treated and control enterprises (clusters) in a job training program but do not discuss the possibility of workers moving across firms.

We make a methodological contribution by developing a framework that clarifies the assumptions and data needed to obtain unbiased estimates of treatment effects in a clustered RCT with movement of service providers across clusters. In our context, this framework explains how treatment effects depend on whether one evaluates, over time, program impacts on the pedagogical practices of the teachers who were in the program schools when the program started or on the pedagogical practices of the teachers in the program schools after turnover has occurred.[2] Both effects are relevant for policy. The first is relevant for policy-makers seeking to raise the skills of particular sets of teachers because, for example, they have a lower level of skills. The second effect is relevant for policymakers seeking to improve teachers' skills in particular schools because, for example, those schools serve disadvantaged students. We show how this latter effect depends not only on the direct effect of the program on participating teachers' skills but also on the indirect effect of the program by its impact on which teachers stay in these schools, which teachers leave these schools, and which teachers move to these schools.

We show that comparisons of teachers in treated and control schools after turnover has taken place will, in general, lead to biased estimates of both types of treatment effects. However, we show that it is possible to estimate an average intent to treat (ITT) effect for the teachers *in treated schools when the program started* if one has a sample of teachers that follows them when they change schools, or using the data of teachers in treated and control schools after turnover has occurred *if* turnover is unrelated to the program. This last result is important because following teachers who change schools and, more generally, following service providers who leave their original cluster, can be difficult,

[2] We use the terms pedagogical skills and pedagogical practices interchangeably, although strictly speaking the former refers to what teachers are capable of doing and the latter refers to what they are observed doing.

which raises the risk of attrition bias in ITT estimates. We also show that it is impossible to estimate any treatment effects for teachers *in the treated schools after turnover occurs*, even if turnover is unrelated to the program, yet the above ITT is a lower bound for the average treatment effect on these teachers.

We estimate the effects on teachers' pedagogical skills of a teacher coaching program implemented in rural multi-grade schools in Peru. Trained coaches visit classrooms and give specific advice to teachers on their pedagogical practices, with customized strategies to improve them. Identification exploits random assignment of 6,207 schools to treatment and control groups when the program expanded in 2016. We randomly selected 182 treated schools and 182 control schools for our evaluation sample. Pedagogical skills were measured in late 2017 (after almost two years of treatment) by observing teacher-student interactions and a broad range of instructional practices.

As in many developing countries, Peru's rural schools have very high rates of teacher turnover;[3] of the teachers in the 364 evaluation sample schools in 2016, about 43% had moved by the start of 2017. Importantly, classroom observation data were collected not only in these 364 schools, but also in many (but not all) of the schools that received the teachers who moved from these schools to other schools between 2016 and 2017.

Our main findings are as follows. If one refers to the pedagogical practices of the teachers who were in the program schools when the program started, we find that offering the program for two years increased teachers' overall pedagogical skills by 0.20 standard deviations (s.d.) of the distribution of those skills. Regarding the pedagogical skills of the teachers who, after turnover has occurred (i.e. at the end of 2017), were teaching in the schools assigned to the program, we find that the effect of the coaching on those skills after two years is at least 0.20 s.d. Turning to specific skills, the largest increases are for lesson planning and, to a lesser extent, encouraging students' critical thinking.

We also estimated the effect of the program on student learning after one year (no data are available for the second year). Combining mathematics and reading test scores, the program increased learning among the Grade 2 students who took the 2016 National Student Evaluation by 0.25 s.d. For Grade 4 students, the effect varies by the number of teachers, with larger effects for schools with only one or two teachers. We also show that the schools most affected in terms of teachers' pedagogical skills are also the schools with

[3] High teacher turnover is common in developing countries: Zeitlin (2021) reports turnover of about 20% per year in Rwanda while Schaffner, Glewwe and Sharma (2021) show that between 18% and 21% of teachers in Nepal changed schools from one year to the next.

the highest increases in test scores. This suggests that changes in pedagogical skills are at least one of the mechanisms linking teacher coaching to student learning.

The estimates we obtain for the effect on pedagogical skills are smaller than those found for coaching programs in developed countries (0.49 s.d. on instructional practices, see Kraft et al., 2018) and this may reflect the scale of the program, and the high rate of teacher turnover in Peru. Yet we also address two unresolved questions regarding the effectiveness of coaching in improving pedagogical skills in developing countries: (i) We show that a program implemented at scale and affected by turnover can still exhibit positive results; and (ii) We show that general pedagogical skills can be improved.

To our knowledge, no previous study has evaluated the effects on pedagogy of a large-scale teacher coaching program in a developing country.[4] Most in-service training programs evaluated in the developing world are small-scale pilots or efficacy trials run by researchers or NGOs (Evans and Popova, 2016). For example, Cilliers et al. (2020) compared the impacts of coaching and centralized teacher training on reading skills in 180 public schools in South Africa. Albornoz et al. (2020) estimated the impact of teacher coaching to improve student learning of science in 70 public schools in Argentina. In contrast, we evaluate a program implemented in 3,795 rural schools in Peru.

The issue of scale is relevant for coaching programs' effectiveness because of two characteristics of this type of in-service training. First, the program's success depends on the supply of qualified coaches. If these skills are scarce, expanding the program likely will reduce its quality, and thus its effectiveness. Second, classroom observation and personalized feedback requires coaches to commute to several schools. This can be costly and can complicate program delivery if scaling-up implies serving schools in very remote areas. This is very likely for rural schools in developing countries, whose teachers often require additional training.

Teacher turnover will not only complicate the identification of program effects, as discussed above, but is also a potential threat to effective coaching programs as it will reduce compliance. Teachers who leave a school while the program is being implemented may not receive the full "dose" of coaching, and program schools that receive new teachers will have staff who are only partially trained. Even if a teacher moves from one

[4] Majerowicz and Montero (2021) estimate the effect on student learning of the same program evaluated in this study. They find large (0.25-0.38 s.d.) and statistically significant effects. We complement these findings by focusing on pedagogical skill as a relevant mechanism linking coaching to student learning.

school with the program to another with the program, switching to another coach may reduce the program's impact, relative to having the same coach for the entire treatment.

We know of only one other study that considered teacher turnover when evaluating a teacher training program. Clare et al. (2010) estimated the effect of a literacy coaching program in 32 elementary schools in Texas. Stressing how such turnover can thwart schools' efforts to improve instruction through teacher training, the authors estimated the program's effect on the reading skills of the students of teachers recruited to replace those who left their school in the first year of the program. They found a positive association between teachers' program participation and their students' reading skills. However, the non-random composition of their sample (recruited teachers in program and non-program schools may not be comparable) casts doubt on the causal interpretation of their results.

Finally, the literature thus far does not provide a clear indication as to whether coaching can improve general pedagogical skills. Most evaluations of coaching programs focus on pedagogy for a specific topic or course. For, example, Albornoz et al. (2020) focused on improving teaching of science, and Cilliers et al. (2020) focused on reading skills. Kraft et al. (2018) highlight a lack of causal evidence on the effect of coaching for subjects other than reading or literacy. Some papers measure the effect of training on teacher time allocation (Bruns et al. 2018) or on using specific types of teaching (Kotze et al. 2019), but not on their teaching skills. The pedagogical skills of public school teachers in developing countries are generally low, and a key policy question is whether coaching can improve a broad set of teaching skills.

The rest of the paper is organized as follows. Section 2 describes the program and explains the evaluation design. Section 3 presents our analytical framework, defines several treatment effects, and explains which ones can be estimated. Sections 4 and 5 present our estimates of the impact of the program on teachers' pedagogical skills and on student learning, respectively. Finally, Section 6 provides concluding remarks, policy implications and suggestions for further research.

## 2. The Coaching Program and its Evaluation Design

In 2010, the Peruvian government initiated coaching programs to improve public primary school teachers' pedagogical practices. As per Ministry of Education guidelines, the local education authority (UGEL) hires coaches for teachers in the schools targeted by the program. Coaches were to be selected from top-performing teachers. Applicants needed

to have a pedagogical college or university degree, at least five years of experience as a primary school teacher, and at least one year of previous experience training or providing support to teachers. Coaches were paid the equivalent of US$1,200 a month, about double the average wage of a teacher at the time. The Ministry of Education set the standards for hiring the coaches (as well as the design of APM in general), but the UGELs selected and hired the coaches. Coaches were trained by UGEL training specialists, who were themselves trained by Ministry of Education officials.

A coach's work consists of several steps. First, the coach meets with the school principal and gathers information about the educational context. Then, the coach attends all teachers' class sessions (one teacher per day) to observe their classroom performance and make an initial diagnostic assessment. The coach uses this assessment to identify the competencies that the teachers must improve and, with each teacher, develops an improvement plan. During the school year, the coach observes eight more of each teacher's class sessions at regular intervals. The program is usually implemented for several consecutive years. After each classroom observation, the coach and the teacher meet to discuss the progress made in terms of the improvement plan. The coach sends monthly and quarterly reports to the UGEL, and to the school principal, on each teacher's progress and on areas for future improvement. At the end of each year, the coach provides a final feedback session for each teacher, collecting his or her impressions of the process. The coach then writes a final report for each teacher on the achievements, actions, and areas requiring further effort, referencing the initial improvement plan.

These programs are a substantial investment by Peru's government, costing over US$ 130 million per year. By 2016, teachers in over 14,000 schools with more than 900,000 students were being coached under several coaching programs. Over 90% of the participating schools are primary schools. For these schools there are three versions of the program: (i) bilingual coaching (for schools where most students speak a Peruvian indigenous language); (ii) monolingual multi-grade coaching (for schools where most students speak Spanish and there are fewer teachers than grades taught – the program may tend to focus on Grade 1 and 2 teachers in schools with more than one teacher);[5] and (iii) monolingual full-teacher coaching (for schools large enough to have one teacher per grade).

This paper evaluates the second type of coaching program. Of Peru's 22,336 rural primary schools, 20,744 are multi-grade, which typically have two teachers and about 30

[5] Although coaches should have given equal time to all teachers, an informal emphasis was given to grade 1 and 2 teachers since the standardized test used to track the education system was given to grade 2 students.

students. In 2016, this program, called *Acompañamiento Pedagogico Multigrado* (APM) in Spanish, was expanded in a way that involved random assignment. All schools that started the program before 2016 continued to participate in APM. Monolingual multi-grade schools that had low scores on Peru's Grade 2 national student evaluation and had not yet participated in APM were randomized into treatment and control groups. Of the 6,207 eligible schools, 3,795 were randomly assigned to the treatment group and started the program in February of 2016 (Peru's school year runs from February to November). The other 2,412 schools, the control group, did not participate in any coaching program in 2016 and 2017. This randomization was stratified at the region (department) level, Peru's highest level of political division (Peru has 26 regions).

A random sub-sample of 364 schools, stratified at the region level, was selected for this study: 182 randomly selected from the 3,795 treated schools, and 182 randomly selected from the 2,421 control schools. Teachers' pedagogical practices were observed in these 364 schools at the end of the 2017 school year. Also, many teachers who had left these 364 schools to go to other schools in 2017 were followed and observed in their new schools. The observers assessed eight pedagogical skills of these teachers (see Table 1). We also construct an overall index by standardizing and then averaging these eight skills.

## 3. Framework and Treatment Effects

Teacher turnover can compromise compliance and introduce spillovers. We evaluate APM in schools where this program had been operating for two years, 2016 and 2017; yet at the end of the first year many teachers in the schools assigned to receive APM moved to schools that did not offer it. Also, some teachers assigned to a non-APM (control) school in the first year moved in the second year to a school that offered APM and so received one year of treatment. From the point of view of schools, in the second year (2017), some APM schools received new teachers with no prior APM coaching, and some non-APM schools received teachers who had been in an APM school in the first year (2016).

Teacher turnover can also introduce new mechanisms through which APM can affect the initial (pre-program) pedagogical skills of the teachers in the treated schools. For example, the program can affect the composition of (initial) pedagogical skills in the APM schools by attracting teachers with higher or lower levels of those skills.

Some structure is needed to account for these phenomena. This section presents the assumptions we impose to use the data that we have to estimate APM's effects on

**Table 1: Description of the Pedagogical Skills on which Teachers Were Assessed**

| Pedagogical Skill | Description |
|---|---|
| Lesson Planning | The session's purpose is stated explicitly, in a way that students can understand. Activities are planned and aligned with the stated purpose. The session is closed referring to its purpose. |
| Time Management | Almost all time is allocated to pedagogical activities. Routines, transitions, and interruptions are well managed. Students know the routines and require little teacher assistance to do them. |
| Promotion of Students' Critical Thinking | The activities promote analysis and reasoning. Most of the questions are open ended and students are given time to delve into them. |
| Promotion of Students' Participation | The teacher succeeds in getting students involved and actively participating, incorporating their opinions, ideas, and interests into the session. Students can influence the class dynamics. |
| Provision of Oral Feedback | The teacher pays attention to the difficulties, doubts, and errors of the students, encouraging them to develop their own answers (through questions or hints), helping them to improve their understanding of the subject and advancing in their learning process. The teacher gathers evidence of the students' progress. |
| Provision of Written Feedback | The teacher assesses the students' work, helping them to see how to achieve what is expected of them. |
| Quality of Relations Between Teacher and Students | Relationships in the classroom are respectful. The class sessions possess a warm environment. |
| Management of Students' Behavior | The teacher employs positive strategies to promote and reinforce good behavior of students, who autoregulate. An environment that promotes learning is facilitated. Bad behavior is very rare. |

teachers' pedagogical skills, accounting for the effects of teacher turnover on compliance, on the emergence of spillovers, and on changes in the composition of teachers' (initial) pedagogical skill within the APM (treatment) and non-APM (control) schools.

### 3.1 A Production Function for Pedagogical Skill

Assume that pedagogical skill, a stock variable, is an increasing function of experience. The effect of experience depends on the teacher (some are better at using experience than others to increase their skills) and whether the school where he or she works offers APM.

The pedagogical skill of teacher $i$ at the end of year $t$, $y_i^t$, is a function of: (i) the skill he or she had at the end of the previous year ($y_i^{t-1}$); (ii) the teacher-specific effect of one year of experience ($\lambda_i$); and (iii) whether his or her school offered APM in year $t$. The presence of APM in the school where teacher $i$ worked in year $t$ is indicated by $T_i^t$ ($T_i^t = 1$ if the school had APM in year $t$ and $T_i^t = 0$ if it did not). This yields:

$$y_i^t = F(y_i^{t-1}, \lambda_i, T_i^t; \gamma_i) \quad (1)$$

where $\gamma_i$ is a set of parameters governing the relation between $y_i^t$ and the three inputs, which can vary over teachers. We make (and later test) the assumption that there are no explicit complementarities between these inputs; this leads to a linear production function:

$$y_i^t = \rho y_i^{t-1} + \lambda_i + \delta_i T_i^t \quad (2)$$

Equation (2) stipulates that, in year *t*, teacher $i$ retains a proportion $\rho$ of the pedagogical skill attained in year *t* - 1 and accumulates additional skill from experience. Also, a teacher can further enhance his or her skill by $\delta_i$ by working in an APM school.

APM was randomly assigned within the evaluation sample schools at the beginning of year 1 (2016), and the evaluation data were collected at the end of year 2 (2017). Therefore, the pedagogical skill of teacher $i$ at the end of year 2, $y_i^2$, can be expressed as:

$$\begin{aligned} y_i^2 &= \rho y_i^1 + \lambda_i + \delta_i T_i^2 = \rho(\rho y_i^0 + \lambda_i + \delta_i T_i^1) + \lambda_i + \delta_i T_i^2 \\ &= \rho^2 y_i^0 + (1+\rho)\lambda_i + \delta_i(\rho T_i^1 + T_i^2) \end{aligned} \quad (3)$$

For simplicity, assume that $\rho = 1$; this is reasonable since the depreciation of skill in one year is likely to be negligible. This implies that: $y_i^2 = y_i^0 + 2\lambda_i + \delta_i(T_i^1 + T_i^2)$. Our objective is not to identify all the parameters of this equation. So, for simplicity, $y_i^0$ and $2\lambda_i$ are combined into a teacher-specific component, denoted by $\theta_i^2$ $(= y_i^0 + 2\lambda_i)$. Thus, teacher *i*'s pedagogical skill at the end of year 2 (end of 2017) can be expressed as:

$$y_i^2 = \theta_i^2 + \delta_i(T_i^1 + T_i^2) \quad (4)$$

The population of interest for this study is the teachers in the 6,207 multi-grade schools involved in the randomized expansion of APM that started in year 1 (2016). The per year average treatment effect (ATE) is defined as $\delta$. That is, $\delta \equiv E[\delta_i]$, where this expectation is over the population of teachers working in the 6,207 multi-grade schools.

### 3.2 Four Types of Teachers

There are (at least) two distinct ways to draw a sample of teachers from the population of 6,207 multi-grade schools. The first is to sample teachers based on the schools where they worked in year 1 (henceforth, Sample 1). The second is to sample teachers based on the

schools where they worked in year 2 (henceforth, Sample 2). These two samples differ because teachers can move to a different school between years 1 (2016) and 2 (2017).

As will be seen in Section 4, the characteristics of the teachers in Sample 1 who worked in APM schools in year 1 are very similar to those of the teachers in Sample 1 who worked in non-APM schools in year 1. This is because the APM program was randomly assigned to those schools in year 1, and all 364 schools (and the teachers in them in year 1) fully complied with their random assignment. Thus, in year 1 all Sample 1 teachers in the APM (treatment) schools received one year of APM coaching and all Sample 1 teachers in the non-APM (control) schools received no APM coaching. Yet, between year 1 and year 2, many Sample 1 teachers moved to a different school. Some teachers in the APM schools in year 1 moved to a non-APM school and, therefore, did not receive APM coaching in year 2. Conversely, some teachers in non-APM schools in year 1 moved to an APM school and thus received one year of coaching, in year 2.

Unlike Sample 1 teachers, Sample 2 teachers working in APM schools in year 2 may not have the same characteristics as Sample 2 teachers working in non-APM schools in that year. This is possible because teacher turnover between years 1 and 2 can change the composition of teachers in a school, and APM may influence this turnover. Note also that all Sample 2 teachers working in APM schools in year 2 received a year of coaching in that year but not necessarily in year 1. Similarly, all Sample 2 teachers working in non-APM schools in year 2 received no coaching in that year, but may have in year 1.

The Angrist, Imbens and Rubin (1996) framework divides the population of interest into *always takers*, who can always obtain the treatment, *never takers*, who can always avoid the treatment, and *compliers*, who comply with their assigned treatment status. Strictly speaking, these classifications are based on behavior, and do not imply any specific assumptions about preferences.

In the APM context, changes in treatment status occur via turnover (teachers switching schools). Part of this turnover may be driven by the presence of the program, but some part may also occur for reasons other than APM. If turnover is in part due to the program, it is reasonable to assume that such teachers have preferences regarding APM. We propose a framework that allows differences in preferences for APM to explain at least some teacher turnover, but we do not want turnover to be explained only by these preferences; teachers may switch schools for reasons completely unrelated to APM.

This requires changing the "traditional" classification of the population. For example, the traditional Angrist, Imbens and Rubin (1996) framework classifies a teacher

moving from an APM school to a non-APM school as a *never taker*. If we assume that this is driven by a strong preference against APM, and ascribe that preference to *never takers*, we exclude the possibility that this move would have occurred in the absence of APM.

To allow for teacher turnover that is unrelated to APM, we divide the population of teachers in the 6,207 multi-grade schools into four groups. First, we divide teachers into those who are relatively indifferent to APM and those with strong preferences for or against it. We further separate teachers in the former group into those who, independently of APM, change schools, whom we call *movers* (M), and those who remain in their schools, whom we call *remainers* (R). We then divide teachers with strong preferences for or against APM into those who like APM, whom we call *likers* (L), and those who dislike APM, whom we call *dislikers* (D).[6] We allow the impact of APM to differ by the type of teacher. Thus, we define $\delta^M, \delta^R, \delta^L$ and $\delta^D$ as the average effects of one year of APM on movers, remainers, likers and dislikers, respectively.

Since all 364 evaluation sample schools followed their random assignment in 2016, all teachers had no choice regarding participation in APM in year 1.[7] We assume that teachers' behavior between years 1 and 2 can be summarized as follows: (i) by definition, all likers assigned to non-APM schools in year 1 move to an APM school in year 2, and all dislikers assigned to APM schools in year 1 move in year 2 to a non-APM school; (ii) all likers assigned to APM schools and all dislikers assigned to non-APM schools do not switch schools between years 1 and 2; (iii) likers and dislikers choose schools in year 2 before the movers, so that movers take whichever teaching positions are available after all likers and dislikers who want to change schools have done so (since movers have no preference between APM and non-APM schools); (iv) the number of teacher positions in APM and non-APM schools is fixed, and there are enough teacher positions available to accommodate the transitions described above; and (v) any teacher transitions in or out of the population of 6,207 multi-grade schools in year 2 do not change the proportions of likers, dislikers, movers and remainers that existed in those schools in year 1.

Comparing our four groups of teachers with the "traditional" classification above, *likers* and *dislikers* are equivalent to *always takers* and *never takers,* respectively, and *remainers* can be classified as *compliers*. The key difference is that the behavior of *movers*

[6] As almost all other studies do, we assume that there are no "defiers". Such teachers would move to a control school in year 2 if they were assigned to an APM school in year 1, or move to an APM school in year 2 if assigned to a control school in year 1, *because* they want to defy their random assignment.
[7] When teachers learned of their random assignment for 2016 it was too late to switch schools in that year.

is consistent with the behavior of any of these three traditional groups. If a mover does not change treatment status after changing schools, he or she would be considered a *complier*. Yet if this teacher had moved from an APM school to a non-APM school he or she would be classified as a *never taker*, and if he or she had moved from a non-APM school to an APM school, he or she would be considered an *always taker*. Moreover, since movers do not take APM into account when changing schools, they always have a probability between 0 and 1 (and never equal to 0 or 1) of moving to an APM (or non-APM) school between years 1 and 2, which is not the case for any of the three "traditional" groups.

### 3.3 Defining Treatment Effects

Before considering whether estimates of treatment effects are biased, the treatment effects one wants to estimate must be defined. We use the standard potential outcomes approach, but given that some teachers change schools between years 1 and 2, and that our data require us to focus on the program's impact after two years, we have four –instead of the standard two – potential outcomes. They are:

$y_{0,0} \equiv$ outcome if not treated in either year
$y_{0,1} \equiv$ outcome if not treated in year 1 but treated in year 2
$y_{1,0} \equiv$ outcome if treated in year 1 but not treated in year 2
$y_{1,1} \equiv$ outcome if treated in both years

Virtually all of the schools in the evaluation sample complied with their random assignment for both years.[8] In contrast, because teachers can move each year, many did not comply with their random assignment in year 2 (although all did so in year 1). Thus, to estimate the impact of two years of APM it is useful to distinguish between treatment effects from the perspective of teachers and from the perspective of schools.

For teachers, all four potential outcomes are possible because they did not have to follow their random assignment in year 2. In contrast, schools followed their random assignment in both years, so from the schools' perspective the only potential outcomes are $y_{0,0}$ and $y_{1,1}$. But this does *not* mean that the teachers in them had only those potential outcomes in year 2: in that year the schools randomly assigned to APM had teachers with

[8] All schools followed their random assignment in 2016, and all but two schools did so in 2017. Two APM schools were reclassified as bilingual schools in 2017 and thus became ineligible for this version of APM. Those two schools are included in the analysis, but excluding them does not change the results (results available from the authors upon request).

potential outcomes $y_{1,1}$ *and* $y_{0,1}$, and schools randomly assigned to the control group had teachers with potential outcomes $y_{0,0}$ *and* $y_{1,0}$. Most importantly, from the perspective of teachers $\theta_i^2$ is fixed and cannot be changed. Yet, from the perspective of schools (average) $\theta_i^2$could change due to teachers who move from year 1 to year 2: likers, who all move to APM schools in year 2 could have an (average) $\theta_i^2$ different from that of the dislikers, who all move to non-APM schools in that year. These two perspectives correspond to the two policy perspectives that were proposed in the Introduction: (i) The effect of APM on the pedagogical practices of the teachers in the APM schools when that program started; and (ii) The effect of APM on the pedagogical practices of the teachers who, at the end of year 2, were in the schools assigned to the APM program.

These four potential outcomes and the two different perspectives lead to several possible definitions of treatment effects. The rest of this subsection defines the treatment effects that we believe are most relevant for evaluating the impact of the APM program, which we then compare to the estimators that we can implement with our data.

***3.3.1 Definition of Treatment Effects from the Perspective of Teachers.*** APM is a program that treats teachers. From the perspective of teachers, we make the standard SUTVA assumption that the program's effect on any teacher's potential outcomes does not depend on whether other teachers participated in the program; it depends only on whether that particular teacher participated.[9] This assumption is implicit in equation (1).

Consider the definition of the average treatment effect (ATE) for teachers. In the most general framework, we can define three different ATEs for teachers:[10]

$\mathrm{ATE}_{1,1} \equiv \mathrm{E}[\mathrm{y}_{1,1}^2 - \mathrm{y}_{0,0}^2]$ (compare treatment in both years relative to no treatment)
$\mathrm{ATE}_{1,0} \equiv \mathrm{E}[\mathrm{y}_{1,0}^2 - \mathrm{y}_{0,0}^2]$ (compare treatment in first year only relative to no treatment)
$\mathrm{ATE}_{0,1} \equiv \mathrm{E}[\mathrm{y}_{0,1}^2 - \mathrm{y}_{0,0}^2]$ (compare treatment in second year only relative to no treatment)

Applying equation (4) to these three treatment effects yields:

[9] This assumption implies that teachers moving from APM to non-APM schools between years 1 and 2 do not influence the pedagogical skills of the teachers in the receiving schools. This assumption is reasonable since APM is a coaching program tailored to the specific needs of each teacher, and in multi-grade schools each teacher teaches grades not taught by other teachers, reducing opportunities to share teaching strategies.
[10] For the y terms, superscripts denote year measured (usually year 2); subscripts denote treatment status.

$$\mathrm{ATE}_{1,1} = 2\bar{\delta} \tag{5a}$$

$$\mathrm{ATE}_{1,0} = \bar{\delta} \tag{5b}$$

$$\mathrm{ATE}_{0,1} = \bar{\delta} \tag{5c}$$

where $\bar{\delta}$ is the impact of 1 year of APM, averaged over all four types of teachers in the population of 6,207 rural multi-grade schools involved in the randomized expansion. Once one knows any one of these three ATEs, the other two are also known. Since the data available measure outcomes after two years, and (as seen below) most of the teachers who were treated were treated for two years, we focus on estimating $\mathrm{ATE}_{1,1}$. Henceforth, we replace the "1,1" subscript with "teachers", to distinguish from ATE from the perspective of schools, and we define ATE from the perspective of teachers as:

$$\mathrm{ATE}_{\mathrm{teachers}} \equiv \mathrm{E}\left[y^2_{1,1} - y^2_{0,0}\right] = 2\bar{\delta} \tag{6}$$

This parameter is the effect of two years of APM on the pedagogical practices, averaged over all teachers in the 6,207 schools. The counterfactual is no treatment in either year.

Next, consider the average treatment effect on the treated (ATT) for teachers. The standard approach defines ATT as $\mathrm{E}[y_1 - y_0 | T = 1]$, where T = 1 denotes a unit (in this case, a teacher) that has been treated. We follow that approach, but this requires clarification about what T = 1 means. We define "T = 1" as whether the teacher is treated in year 2, since teachers had no choice in year 1. Intuitively, ATT is well defined only if at least some individuals have a choice regarding whether they are treated, and the only choice in the APM context is treatment in year 2, which we denote by $T^2$.[11] The teachers for whom $T^2 = 1$ are: all likers, no dislikers, a random sample of remainers, and movers who happen to be in treated schools in year 2 (which is a random sample of movers).

There are two different definitions of ATT for teachers, depending on how the counterfactual is defined. The first, which can be denoted as $\mathrm{ATT}_1$, specifies the counterfactual as not being treated in year 2 while leaving treatment status in year 1 as teachers' actual random assignment, which we denote by $y^2_{T^2=0}$. This ATT is defined as:

[11] The only other plausible definition of ATT would be what teachers would have done if they had also had a choice in year 1. Then all likers would get two years, and all dislikers zero years, of treatment, and remainers and movers would behave the same as they do in the case where teachers have a choice only in year 2.

$$\begin{aligned} \text{ATT}_1 \equiv\ & \text{E}\left[y^2 - y^2_{T^2=0} \middle| T^2 = 1\right] \\ =\ & \delta^R p^R + \delta^L (p^L/\tau) + \delta^M (1 - p^R - p^L/\tau) \end{aligned} \tag{7}$$

where $\tau$ is the proportion of teaching positions in the population of 6,207 multi-grade schools that are in the APM schools, and $p^R$, $p^L$, $p^D$ and $p^M$ are the proportions of the teachers in these 6,207 schools who are remainers, likers, dislikers and movers, respectively. As explained in subsection 3.2, we assume that these proportions do not change between years 1 and 2. See Online Appendix 1 for the derivation of equation (7).

The parameter given in (7) is the effect of one year of APM on the pedagogical practices of the teachers who were treated in year 2. The counterfactual is that these teachers were not treated in year 2, but treatment status in year 1 is actual treatment in that year. The intuition for equation (7) is the following. Remainers who are treated in year 1 accept another year of treatment. All likers choose one more year of treatment, and since they all move to APM schools their proportion in APM schools in year 2 increases by a factor of $1/\tau$. Finally, the proportion of teachers in APM schools in year 2 who are movers, who "randomly chose" schools in year 2, is $(1 - p^R - p^L/\tau)$, that is the proportion of teaching positions in those schools that are not taken by remainers or likers.[12]

The second possible counterfactual for ATT is one where APM does not exist, so the counterfactual is $y^2_{0,0}$. The only change to the definition for $\text{ATT}_1$ in equation (7) is that the counterfactual $y^2_{T^2=0}$ is replaced by $y^2_{0,0}$. We call this $\text{ATT}_2$; it is the effect of APM after two years on the pedagogical practices of the teachers who are treated in year 2:

$$\begin{aligned} \text{ATT}_2 \equiv\ & \text{E}\left[y^2 - y^2_{0,0} \middle| T^2 = 1\right] \\ =\ & 2\delta^R p^R + (1+\tau)\delta^L (p^L/\tau) + (1+\tau)\delta^M (1 - p^R - p^L/\tau) \end{aligned} \tag{8}$$

The derivations for the second line of equation (8) are given in Online Appendix 1.

$\text{ATT}_2$ is also very intuitive. As with equation (7), the proportions of the teachers in APM schools in year 2 are $p^R$ remainers, $p^L/\tau$ likers, and $(1 - p^R - p^L/\tau)$ movers. All remainers received two years of treatment. The average liker received $1+\tau$ years of treatment, which is an average over the proportion of likers who were in APM schools in year 1 $(1-\tau)$ and so received only one year of treatment and the proportion of likers who

[12] As explained in subsection 3.2, we assume that there are enough spaces in the APM schools for all likers, $1 - p^R - p^L/\tau \leq 1$, and there are enough spaces in non-APM schools for all dislikers, $1 - p^R - p^D/(1-\tau) \leq 1$.

were in treated schools in year 1 ($\tau$) and so received two years of treatment: $1\text{-}\tau + 2\tau = 1+\tau$. The average mover in a treated school in year 2 also received $1+\tau$ years of treatment; unlike likers, movers in APM schools in year 2 are there by random chance, yet as with likers they are all treated in year 2, and $\tau$ is the proportion treated in year 1.

Next, consider the intention to treat (ITT) effect from the perspective of teachers. In general, ITT is defined as the difference in the observed outcome when one group is randomly assigned to the program and the other is randomly assigned to a control group. From the perspective of teachers, this is the random assignment that occurred in year 1. Thus, ITT for teachers can be defined as this difference in *observed* y's in year 2:

$$\begin{aligned} \text{ITT} &\equiv \text{E}[y^2|R^1 = 1] - \text{E}[y^2|R^1 = 0] \\ &= \bar{\delta} + \delta^R p^R \end{aligned} \tag{9}$$

where $R^1$ refers to random assignment in year 1 and $\bar{\delta}$ is the population average treatment effect ($\delta^{\text{R}}\text{p}^{\text{R}} + \delta^{\text{L}}\text{p}^{\text{L}} + \delta^{\text{D}}\text{p}^{\text{D}} + \delta^{\text{M}}\text{p}^{\text{M}}$). See Online Appendix 1 for this derivation.

Similar to the $\text{ATE}_{\text{teachers}}$, the parameter given in (9) is a two-year effect on the teachers originally assigned to treatment. Unlike $\text{ATE}_{\text{teachers}}$, however, is that it is the two-year effect of *assignment* to the program in year 1. The counterfactual is assignment to the control group (non-APM schools) in year 1. Remainers are treated for two years because they comply with their random assignment for both years. Likers and dislikers move to the schools that they prefer in the second year, so the effect of their random assignment in the first year lasts for only one year, and thus those randomly assigned to APM schools in year 1 get one more year of treatment than those randomly assigned to non-APM schools in year 1. Finally, movers move randomly in year 2, regardless of their random assignment in year 1, which implies that they also are affected by their random assignment only in year 1, so those randomly assigned to APM schools in year 1 get one more year of treatment than those randomly assigned to non-APM schools in year 1.

Finally, consider the local average treatment effect (LATE) from the perspective of teachers. To define LATE, two new variables need to be defined:

$$P_0 = \text{ value of } T^2 \text{ if } R^1 = 0 \tag{10a}$$

$$P_1 = \text{ value of } T^2 \text{ if } R^1 = 1 \tag{10b}$$

LATE is defined as the effect of two years of APM on those teachers who, in year 2, are *certain* to comply with their random assignment in year 1, which is all teachers for whom $E[P_0] = 0$ and $E[P_1] = 1$. This includes all remainers, but excludes the other three types of teachers; even though some movers randomly "mimic" the behavior of remainers (are in the same type of school in year 1 as in year 2), this happens by chance and is unrelated to their random assignment in year 1. Remainers get either two years or no years of treatment, because their treatment status does not change from year 1 to year 2. Thus, this definition of LATE from the perspective of teachers, denoted by $LATE_1$, is:

$$\begin{aligned} \text{LATE}_1 \equiv\ & \text{E}[y^2_{1,1} - y^2_{0,0} |\, \text{E}[P_0] = 0, \text{E}[P_1] = 1] \\ =\ & 2\delta^R \end{aligned} \tag{11}$$

That is, $LATE_1$ is the impact of two years of being in an APM school on the remainers, who are analogous to the compliers in Angrist, Imbens and Rubin (1996).

There is another version of LATE. It incorporates the fact that *all* teachers comply with their random assignment in year 1, but random assignment to an APM school in that year causes only remainers to be treated for two years; for the other three types of teachers random assignment to an APM school leads to only one more year of treatment. (Some likers and movers are also treated for two years, but they would have been treated in year 2 even if they had been randomly assigned to the control group in year 1.) This treatment effect, which we denote by $LATE_2$, classifies all teachers as compliers, but this weighted average of treatment effects for the four types of teachers gives "double weight" to remainers since their random assignment to APM schools leads to two years of treatment, while random assignment of all other teachers yields only one more year of treatment.

More specifically, we define $LATE_2$ as:

$$\begin{aligned} \text{LATE}_2 \equiv\ & (2\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M)/(2p^R + p^L + p^D + p^M) \\ =\ & \text{ITT}/(1 + p^R) \end{aligned} \tag{12}$$

Two aspects of $LATE_2$ are worth noting. First, as shown in equation (12), it equals ITT divided by the sum of the weights. Second, strictly speaking it is not a measure of the impact of two years of the program; instead, it is a weighted average of the one year impacts, with "double weight" given to the impact on remainers.

We conclude by comparing the relative sizes of these different treatment effects, under the assumption that the population contains all four types of teachers (all p terms are > 0) and that the four average treatment effects (δ terms) are all ≥ 0. First, it is not possible to make any comparisons for $LATE_1$, since $\delta^R$ can vary between zero and some value larger than that of all the other δ terms. Second, $ATE_{teachers} \geq ITT$ since $ATE_{teachers} - ITT = \delta^L p^L + \delta^D p^D + \delta^M p^M$; this inequality is strict if any of these three δ terms is > 0. This inequality is intuitive because the counterfactual for $ATE_{teachers}$ is no treatment at all, while the counterfactual for ITT in year 2 includes both dislikers and movers who were treated in year 1 because they were *assigned* to an APM school in that year.

Third, $ATT_2 \geq ATT_1$ since $ATT_2 - ATT_1 = \delta^R p^R + \delta^L p^L + \tau\, \delta^M(1 - p^R - p^L/\tau)$, which cannot be negative, and which is strictly positive if any of these three δ terms is > 0. This is also intuitive because $ATT_2$ includes the effect of the first year of APM because its counterfactual is no treatment at all, so in effect $ATT_2$ measures the effect of two years of treatment, while $ATT_1$ measures the effect of only one year of treatment (treated in year 2). Fourth, clearly $LATE_2 < ITT$, since $LATE_2 = ITT/(1 + p^R)$. The intuition here is that $LATE_2$ measures the (weighted) average of one year of treatment, while ITT measures the effect, after two years, of assignment to the program, and all teachers complied with their assignment in year 1.. Finally, it is not possible to determine whether the two ATT treatment effects are larger or smaller than $ATE_{teachers}$, ITT or $LATE_2$ because the ATT expressions include terms that are divided by τ, and these two treatment effects could be very large if τ is close to zero. Alternatively, the two ATT terms do not contain $\delta^D$, while $ATE_{teachers}$, ITT or $LATE_2$ all contain $\delta^D$, so the ATT terms could be smaller than the latter treatment effects if $\delta^D$ were much larger than the other δ terms.

***3.3.2. Definition of Treatment Effects from the Perspective of Schools.*** We can also define treatment effects from the perspective of schools. Aside from the two schools unable to continue with APM in year 2, this scenario is easier since all schools follow their random assignment, which did not change between years 1 and 2. We define the counter-factual as what would have happened had there been no APM program at all, as we did above from the perspective of teachers for $ATE_{teachers}$, and for $ATT_2$.[13] Note that, although these treatment effects are from the perspective of schools, we give teachers equal weight

[13] The counterfactual for $ATT_1$ allows some teachers' treatment status in year 1 to be different from their treatment status in year 2. But schools cannot change their treatment status from year 1 to year 2, so this is not a possible counterfactual from the perspective of schools. Note also that since schools do not change their treatment status from year 1 to year 2, ATT, and LATE, from the perspective of schools equals $ATE_{schools}$, or alternatively ATT, and LATE, are not defined since, in effect, schools had no choice.

when defining them (although this has little effect since the proportion of teaching positions in treated schools is similar to the proportion of schools that are treated schools).

We define $ATE_{schools}$ as the effect of the program after two years on the staff of the average school. Thus we have (see the Online Appendix 1 for details):

$$\begin{aligned} \text{ATE}_{\text{schools}} &\equiv \text{E}[y^2|T^2 = 1] - \text{E}[y^2_{0,0}] \\ &= 2\delta^R p^R + (1+\tau)\delta^L(p^L/\tau) + (1+\tau)\delta^M(1 - p^R - p^L/\tau) \\ &+ \theta_{2,L}p^L(1-\tau)/\tau - \theta_{2,D}p^D + \theta_{2,M}(1 - p^R - p^L/\tau - p^M) \end{aligned} \tag{13}$$

The intuition for this treatment effect is as follows. The second line in (13) is the impact of the treatment on the teachers' pedagogical skills, while the third line is a composition effect. The second line is identical to $ATT_2$, as one would expect because the counter-factuals are the same; thus the intuition is also the same. Turning to the composition effect, there is no effect for remainers because they do not switch schools. Yet there are effects for likers, dislikers and (with one exception) movers. The APM schools gain $p^L[(1-\tau)/\tau]$ likers and lose all ($p^D$) dislikers, which changes the overall skill composition in those schools by $\theta^{2,L}p^L(1-\tau)/\tau - \theta^{2,D}p^D$. Finally, the composition also changes for movers, and whether this results in more or fewer movers in APM schools relative to the counterfactual depends on the relative sizes of $(1 - p^R - p^L/\tau)$, the proportion of movers in the APM schools, and $p^M$, the proportion of movers in those schools if APM did not exist.[14] These composition effects rule out unambiguous comparisons of $ATE_{schools}$ with all the treatment effects from the perspective of teachers because the θ terms could reverse any relationships that may hold by comparing only the δ terms.

***3.3.3. No Likers or Dislikers.*** The treatment effects defined above become much simpler if there are no likers or dislikers in the population of teachers. Here we briefly discuss how the treatment effects defined above change under this assumption.

If there are no likers or dislikers, so that $p^L = 0$ and $p^D = 0$, then our parameters of interest simplify as follows (note that $p^R + p^M = 1$):

[14] These two terms would cancel out if $p^L/\tau = p^L + p^D$, in which case $(1 - p^R - p^L/\tau) = p^M$. The intuition is that $p^M = 1 - p^R - p^L - p^D$, so $(1 - p^R - p^L/\tau) = p^M$ implies that $p^L + p^D = p^L/\tau$, and thus $p^D = p^L(1/\tau - 1) = p^L((1-\tau)/\tau)$, so $p^D\tau = p^L(1-\tau)$. When $p^D\tau = p^L(1-\tau)$, the proportion of dislikers leaving APM schools equals the proportion of likers moving into APM schools, so the proportion of movers is unchanged.

$$\mathrm{ATE}_{\mathrm{teachers}} \equiv \mathrm{E}[y^2_{1,1} - y^2_{0,0}] = 2\bar{\delta} = 2\delta^R p^R + 2\delta^M p^M \quad (6')$$

$$\mathrm{ATT}_1 \equiv \mathrm{E}\left[y^2_{T^2=1} - y^2_{T^2=0} \middle| T^2 = 1\right] = \delta^R p^R + \delta^M(1 - p^R) = \delta^R p^R + \delta^M p^M = \bar{\delta} \quad (7')$$

$$\begin{aligned}\mathrm{ATT}_2 \equiv \mathrm{E}\left[y^2_{T^2=1} - y^2_{0,0} \middle| T^2 = 1\right] &= 2\delta^R p^R + (1+\tau)\delta^M(1 - p^R) \\ &= 2\delta^R p^R + (1+\tau)\delta^M p^M\end{aligned} \quad (8')$$

$$\mathrm{ITT} \equiv \mathrm{E}[y^2 | R^1 = 1] - \mathrm{E}[y^2 | R^1 = 0] = \bar{\delta} + \delta^R p^R = 2\delta^R p^R + \delta^M p^M \quad (9')$$

$$\mathrm{LATE}_1 = 2\delta^R p^R \quad (11')$$

$$\begin{aligned}\mathrm{LATE}_2 &= (2\delta^R p^R + \delta^M p^M)/(2p^R + p^M) \\ &= \mathrm{ITT}/(1 + p^R)\end{aligned} \quad (12')$$

$$\begin{aligned}\mathrm{ATE}_{\mathrm{schools}} &\equiv \mathrm{E}[y^2_{T^2=1} | T^2 = 1] - \mathrm{E}[y^2_{0,0}] \\ &= 2\delta^R p^R + \delta^M(1+\tau)(1 - p^R) + \theta_{2,M}(1 - p^R - p^M) \\ &= 2\delta^R p^R + (1+\tau)\delta^M p^M\end{aligned} \quad (13')$$

The following relations hold if there are no likers and dislikers:

$$\mathrm{ATE}_{\mathrm{teachers}} \geq \mathrm{ATT}_2 = \mathrm{ATE}_{\mathrm{schools}} \geq \mathrm{ITT} \geq \mathrm{ATT}_1 \quad (14)$$

$$\mathrm{ATE}_{\mathrm{teachers}} = 2\mathrm{ATT}_1 \quad (15)$$

$$\mathrm{LATE}_2 < \mathrm{ITT} \quad (16)$$

Notice that the composition effect for $\mathrm{ATE}_{\mathrm{schools}}$ vanishes when there are no likers or dislikers, so that $\mathrm{ATT}_2 = \mathrm{ATE}_{\mathrm{schools}}$. Note also that $\mathrm{ATT}_2 \geq \mathrm{ITT}$, since the only difference is that the counterfactual for $\mathrm{ATT}_2$ is that no treatment exists, while the counterfactual for ITT is assignment to the control group in year 1 (and movers may move in year 2). In fact, $\mathrm{ATT}_2 - \mathrm{ITT} = \tau\delta^M p^M$, since a proportion $\tau$ of movers assigned to the control group in year 1 are treated in year 2 (which is not part of the effect of being assigned to APM in year 1).

### 3.4. Estimates in Two Samples

Having defined seven parameters that we may want to estimate, we next discuss which ones can be estimated with the data that are available.

***3.4.1. Estimates in Sample 1.*** Using Sample 1 teachers, we can regress the pedagogical skill measured at the end of year 2 on an intercept and these teachers' treatment status in year 1:

$$y_i^2 = \alpha_1 + \beta_1 T_i^1 + \varepsilon_{1i} \tag{17}$$

The OLS estimate of $\beta_1$, denoted by $\hat{\beta}_{1,OLS}$, estimates $\mathrm{E}[y_i^2|T_i^1 = 1] - \mathrm{E}[y_i^2|T_i^1 = 0]$. If we decompose this difference using equation (4) and $\mathrm{p}^{\mathrm{M}}$, $\mathrm{p}^{\mathrm{R}}$, $\mathrm{p}^{\mathrm{L}}$ and $\mathrm{p}^{\mathrm{D}}$, we have:

$$\begin{aligned}\mathrm{E}[y_i^2|T_i^1 = 1] &= [\theta^{2,M}\mathrm{p}^{\mathrm{M}} + \theta^{2,R}\mathrm{p}^{\mathrm{R}} + \theta^{2,L}\mathrm{p}^{\mathrm{L}} + \theta^{2,D}\mathrm{p}^{\mathrm{D}}] \\ &+ [\delta^M \mathrm{p}^{\mathrm{M}}(1 + 1 - \mathrm{p}^{\mathrm{R}} - \mathrm{p}^{\mathrm{L}}/\tau) + \delta^R \mathrm{p}^{\mathrm{R}}(1 + 1) + \delta^L \mathrm{p}^{\mathrm{L}}(1 + 1) + \delta^D \mathrm{p}^{\mathrm{D}}(1 + 0)]\end{aligned} \tag{18}$$

and

$$\begin{aligned}\mathrm{E}[y_i^2|T_i^1 = 0] &= [\theta^{2,M}\mathrm{p}^{\mathrm{M}} + \theta^{2,R}\mathrm{p}^{\mathrm{R}} + \theta^{2,L}\mathrm{p}^{\mathrm{L}} + \theta^{2,D}\mathrm{p}^{\mathrm{D}}] \\ &+ [\delta^M \mathrm{p}^{\mathrm{M}}(0 + 1 - \mathrm{p}^{\mathrm{R}} - \mathrm{p}^{\mathrm{L}}/\tau) + \delta^R \mathrm{p}^{\mathrm{R}}(0 + 0) + \delta^L \mathrm{p}^{\mathrm{L}}(0 + 1) + \delta^D \mathrm{p}^{\mathrm{D}}(0 + 0)]\end{aligned} \tag{19}$$

Thus, when we run an OLS regression of the pedagogical skill on a constant term and teachers' treatment status in year 1, we estimate $\hat{\beta}_{1,OLS}$:

$$\begin{aligned}\mathrm{E}[y_i^2|T_i^1 = 1] - E[y_i^2|T_i^1 = 0] &= \delta^M \mathrm{p}^{\mathrm{M}} + 2\delta^R \mathrm{p}^{\mathrm{R}} + \delta^L \mathrm{p}^{\mathrm{L}} + \delta^D \mathrm{p}^{\mathrm{D}} \\ &= \overline{\delta_1} + \delta^R \mathrm{p}^{\mathrm{R}}\end{aligned} \tag{20}$$

which is the expression for ITT defined in equation (8). Thus, we can consistently estimate the ITT parameter by running an OLS regression using sample 1 teachers.

Estimation of an ITT parameter allows one to (mechanically) obtain an instrumental variable estimate of the form $\hat{\beta}_{1,OLS}/(1 + \mathrm{p}^{\mathrm{R}})$ by regressing $y_i^2$ on predicted years of treatment, instrumented by random assignment in year 1 ($R_i^1$). This is the $\mathrm{LATE}_2$ parameter defined above, so we can consistently estimate $\mathrm{LATE}_2$ using an IV approach.

None of the other five parameters in subsection 3.3 can be estimated using the Sample 1 data.

***3.4.2 Estimates in Sample 2.*** Now consider estimation using Sample 2, which is the sample of the teachers who were in the treatment (APM) and control (non-APM) schools in year 2. We can regress the pedagogical skill measured at the end of year 2 on an intercept and these teachers' treatment status in year 2. This is an RCT-based estimate of the impact of an education intervention randomly assigned at the school level that

defines treatment by the school a teacher is in when the outcome is measured, ignoring teacher turnover. The equation estimated is:

$$y_i^2 = \alpha_2 + \beta_2 T_i^2 + \varepsilon_{2i} \tag{21}$$

The OLS estimate of $\beta_2$, denoted by $\hat{\beta}_{2,OLS}$, estimates $\mathrm{E}[y_i^2|T_i^2 = 1] - \mathrm{E}[y_i^2|T_i^2 = 0]$. Applying equation (4) yields:

$$\begin{aligned}\mathrm{E}[y_i^2|T_i^2 = 1] &= \mathrm{E}[\theta_i^2|T_i^2 = 1] + \mathrm{E}[\delta_i(T_i^1 + T_i^2)|T_i^2 = 1] \\ &= \theta^{2,M}(1 - \mathrm{p^R} - \mathrm{p^L}/\tau) + \theta^{2,R}\mathrm{p^R} + \theta^{2,L}\mathrm{p^L}/\tau + \\ &\delta^M(1 - \mathrm{p^R} - \mathrm{p^L}/\tau)[\tau + 1] + 2\delta^R\mathrm{p^R} + \delta^L\mathrm{p^L}/\tau[\tau + 1]\end{aligned} \tag{22}$$

We have an analogous equation for $E[y_i^2|T_i^2 = 0]$.

$$\begin{aligned}\mathrm{E}[y_i^2|T_i^2 = 0] &= \theta^{2,M}\left[1 - \mathrm{p^R} - \tfrac{\mathrm{p^D}}{1-\tau}\right] + \theta^{2,R}\mathrm{p^R} + \tfrac{\theta^{2,D}\mathrm{p^D}}{1-\tau} \\ &+\delta^M[1 - \mathrm{p^R} - \mathrm{p^D}/(1-\tau)](\tau + 0) + (\tau + 0)\delta^D\mathrm{p^D}/(1-\tau)\end{aligned} \tag{23}$$

An OLS regression using Sample 2 teachers of their average pedagogical skill on their treatment status in year 2 would therefore estimate:

$$\begin{aligned}&\mathrm{E}[y_i^2|T_i^2 = 1] - \mathrm{E}[y_i^2|T_i^2 = 0] \\ &= \theta^{2,M}\left(\frac{\mathrm{p^D}}{1-\tau} - \frac{\mathrm{p^L}}{\tau}\right) + \frac{\theta^{2,L}\mathrm{p^L}}{\tau} - \frac{\theta^{2,D}\mathrm{p^D}}{1-\tau} \\ &+\delta^M(1 - \mathrm{p^R} - (1+\tau)\mathrm{p^L}/\tau + \tau\mathrm{p^D}/(1-\tau)) + 2\delta^R\mathrm{p^R} + (\tau+1)\delta^L\mathrm{p^L}/\tau - \tau\delta^D\mathrm{p^D}/(1-\tau)\end{aligned} \tag{24}$$

Comparing (24) with the parameters of subsection 3.3, the OLS regression with Sample 2 teachers does not provide unbiased estimates of any of those seven parameters. In particular, comparison with $\mathrm{ATE_{schools}}$ reveals the following bias:

$$\begin{aligned}&\mathrm{ATE_{schools}} - (\mathrm{E}[y_i^2|T_i^2 = 1] - \mathrm{E}[y_i^2|T_i^2 = 0]) \\ &= \theta^{2,M}(1 - \mathrm{p^R} - \mathrm{p^D}/(1-\tau) - \mathrm{p^M}) - \theta^{2,L}(\mathrm{p^L}) + \theta^{2,D}(\tau\mathrm{p^D}/(1-\tau)) + \\ &+\delta^M\tau(1 - \mathrm{p^R} - \mathrm{p^D}/(1-\tau)) + \delta^D\tau\mathrm{p^D}/(1-\tau)\end{aligned} \tag{25}$$

This bias comes from the spillover effect that the program has on the control schools. In particular, the program changes the proportion of three of the four types of teachers in non-APMs. This is captured by the terms in parentheses accompanying $\theta^{2,M}$, $\theta^{2,L}$ and $\theta^{2,D}$ in the second line of (25). These terms correspond to the difference between the proportions of movers, likers and dislikers effectively observed in non-APMs in year 2 and the original (population) proportions of these types of teachers. In addition, non-APMs receive movers and dislikers, some of whom were treated during the first year of the program. This also introduces a bias which is captured in the third line of (25). The sign of the bias in equation (25) is ambiguous.[15]

If we assume no likers or dislikers ($\mathrm{p^L} = \mathrm{p^D} = 0$), equation (24) simplifies to:

$$E[y_i^2|T_i^2 = 1] - E[y_i^2|T_i^2 = 0] = \\ = 2\delta^R \mathrm{p^R} + \delta^M(1 - \mathrm{p^R}) = 2\delta^R \mathrm{p^R} + \delta^M \mathrm{p^M} = \bar{\delta} + \delta^R \mathrm{p^R} \qquad (24')$$

which corresponds to ITT. Thus, in a scenario of no likers or dislikers, $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ should be equivalent as they both estimate the same parameter, which corresponds to ITT.

A final useful result regarding estimates from Sample 2 is that one could regress, for Sample 2 teachers, pedagogical skills at the end of year 2 on those teachers' treatment status in year 1, which could differ from their treatment status in year 2 for the Sample 2 teachers who changed schools. This would be estimation of equation (17) using Sample 2 teachers instead of Sample 1 teachers, and would require data on which Sample 2 teachers switched schools between years 1 and 2 and, for those who switched, on the treatment status of the schools where they were teaching in year 1. If such data are available, this regression would provide an unbiased estimate of the ITT (see Online Appendix 1). This would be feasible if accurate administrative data are available on the schools where teachers taught in year 1, or if teachers can accurately recall the programs that were offered in the schools where they worked in that year.

[15] First, the terms $((1 - \mathrm{p^R} - \mathrm{p^D}/(1 - \tau) - \mathrm{p^M})$ and $-\theta^{2,L}$ are both < 0. Second, the $\theta^{2,D}$ and $\delta^D$ terms are both > 0. Finally, the sign of $(1 - \mathrm{p^R} - \mathrm{p^D}/(1 - \tau))$ is ambiguous.

# 4. The Effect of APM Coaching on Pedagogical Practices

### 4.1 Fieldwork Results: Attrition and Balance

The evaluation sample contains 364 schools from the 6,207 involved in the randomized expansion, 182 of which were randomly selected from the 3,975 schools randomly assigned to APM and 182 of which were randomly selected from the 2,412 schools randomly not assigned to APM. The goal of the fieldwork was to observe, in the third quarter of 2017, the pedagogical practices of the teachers who: (i) had worked in one of the 364 evaluation sample schools in 2016 (Sample 1); and (ii) worked in an evaluation sample school in 2017 (Sample 2). The former required visiting schools not in the evaluation sample because many Sample 1 teachers changed schools between 2016 and 2017.

**Table 2: Attrition of Sample 1 and Sample 2 Teachers and Evaluation Sample Schools in Year 2 (2017)**

| | **Sample 1 teachers** | | | **Sample 2 teachers** | | | **Evaluation sample schools** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment (1) | Control (2) | Total (3) | Treatment (4) | Control (5) | Total (6) | Treatment (7) | Control (8) | Total (9) |
| **Original (2016)** | 321 | 341 | 662 | 355 | 384 | 739 | 182 | 182 | 364 |
| **Observed (in 2017)** | 219 | 236 | 455 | 299 | 341 | 640 | 166 | 174 | 340 |
| **Attrition rate (%)** | 0.318 | 0.301 | 0.312 | 0.158 | 0.112 | 0.134 | 0.088 | 0.044 | 0.066 |
| **Difference in attrition rates** | 0.017 (0.036) | | | 0.046* (0.025) | | | 0.044* (0.026) | | |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

It was not possible to observe the pedagogical practices of all Sample 1 teachers (see the first three columns in Table 2). In fact, attrition in Sample 1 is high. This is mainly due to outdated information on the location of teachers at the time the fieldwork was planned (around March of 2017, the beginning of Peru's school year). The teacher location information at that time indicated that 406 schools needed to be visited, including 104 that were not one of the 364 evaluation sample schools, to observe all Sample 1 teachers who were still teaching. During fieldwork, 91.6% (372) of these 406 schools were visited (34 schools in hard-to-reach areas could not be visited), but the outdated information often led to situations where the teachers were not there; they were working in other schools, and by the time this was discovered it was logistically impossible to go to the schools where those teachers were actually working. As seen in Table 2, only 68.8%

(455 out of 662) of the original Sample 1 teachers were observed in 2017. Of the 207 unobserved Sample 1 teachers, 50 (7.6% of the 662) were no longer teaching in public schools, 28 (4.2%) were in one of the 34 schools that were not visited, and 129 (19.5%) were working in a public school that was not in the planned sample of 406 schools. Turning to Sample 2 teachers (those in the 364 evaluation sample schools in year 2), 86.6% (640 out of 739) were observed in year 2 (see columns (4) - (6) of Table 2). In this sample, attrition is mainly due to the 24 evaluation sample schools located in hard-to-reach areas that could not be visited in year 2 (see columns (7) – (9) of Table 2).

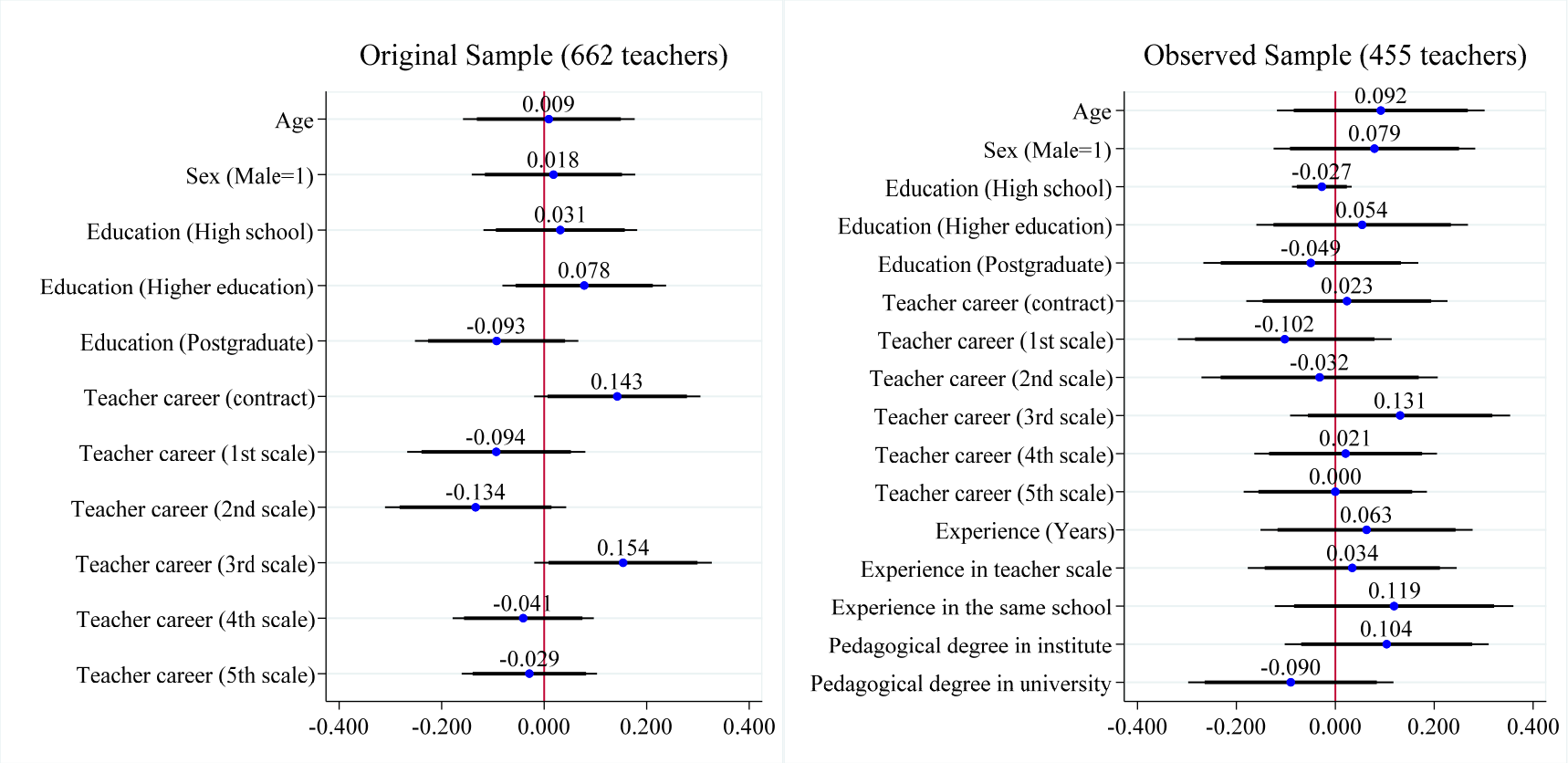
Non-random attrition could lead to biased estimates, especially estimates using Sample 1, given its high rate of attrition. Yet if the average characteristics of the missing teachers are similar for treatment and control teachers, which for Sample 1 would be the case if the data (on where teachers who moved were working) were outdated primarily due to random factors, then this attrition will not yield biased estimates. To check for possible bias, we do two things. First, we compare the attrition rates of the treatment and control groups. Table 2 shows little or no evidence that the rate of attrition was related to the treatment status of teachers or schools. Second, we compare observable characteristics of (non-attrited) schools and teachers belonging to the treatment and control groups.

Random assignment to the program in 2016 should ensure that, before any attrition occurred, the teacher characteristics were balanced for the teachers working in the APM and non-APM schools in that year (Sample 1 teachers). Random assignment should also ensure that the baseline characteristics of the 364 schools in the evaluation sample are balanced. If attrition is random, teacher characteristics should be similar between the teachers working in APM and non-APM schools in 2016 who remained in the subsample of Sample 1 teachers who were observed in 2017 (the 455 teachers in Table 2).

Figures 1 and 2 show that the treatment and control groups are similar in terms of: (i) observed characteristics of the original 662 Sample 1 teachers in year 1 (2016); (ii) observed characteristics of the subsample of 455 Sample 1 teachers who remained in the sample in year 2 (2017); (iii) observed characteristics of the original 364 evaluation sample schools in year 1 (2016); and (iv) observed characteristics of the subsample of 340 schools visited in year 2 (2017). Importantly, none of the (standardized) differences is very large, and none is statistically significant at the 5% level.[16]

[16] The appendix presents additional evidence that attrition is uncorrelated with treatment assignment. Table A3.1 shows that teachers' pre-treatment characteristics do not predict being assigned to the treatment group. Table A3.2 shows that assignment to the treatment group does not predict being observed at the end of 2017.

**Figure 1**
**Balance in Teacher Characteristics for the Original and Observed in Year 2 Teachers Who Worked in an Evaluation Sample School in 2016 (Sample 1)**

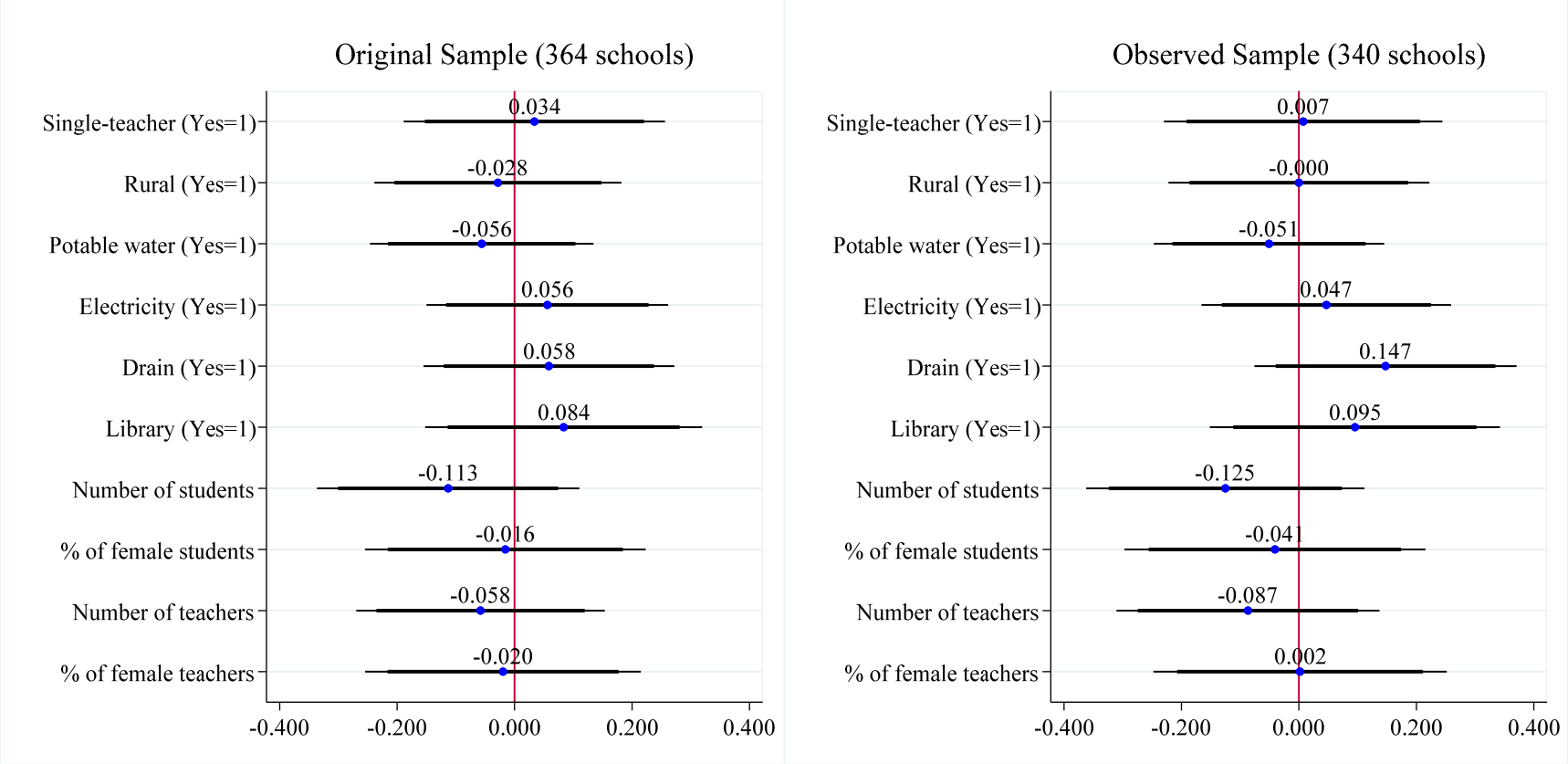



All regressions include UGEL fixed effects. Standard errors clustered at the school level.

Estimates indicate differences in the standardized characteristics of control and treatment groups. Thick and thin lines indicate 90% and 95% confidence intervals, respectively.

We do not present the differences in teacher experience and pedagogical degree for the original sample because we do not have information on those variables for the teachers that were not observed at the end of year 2.

**Figure 2**
**Balance in School Characteristics in the Original and Observed Evaluation Sample Schools**




All regressions include UGEL fixed effects.

Estimates indicate differences in the standardized characteristics of control and treatment groups. Thick and thin lines indicate 90% and 95% confidence intervals, respectively.

We do not compare Sample 2 teacher characteristics between APM and non-APM schools to check for balance at baseline because random assignment in 2016 (year 1) does not ensure that teacher characteristics are balanced across these two groups of schools in 2017 (year 2). In particular, if the program affected the composition of teacher characteristics in APM and non-APM schools in year 2, the characteristics of Sample 2 teachers will be correlated with the treatment status of the schools where they worked in year 2.

### 4.2 Estimating the Proportions of Each Type of Teacher

The data indicate which Sample 1 teachers stayed in their same school between year 1 (2016) and year 2 (2017), and which moved to a different school between these years. The data also show whether the new school hosting those who moved offered APM coaching in year 2. Similarly, the data show which Sample 2 teachers were, in year 2, in the same school as in year 1, and which came from a different school and, for the latter, whether the school from which they came offered APM coaching in year 1.

Tables 3 and 4 present this information for Sample 1 and Sample 2 teachers, respectively. Recall that the evaluation sample is a subsample of a larger group of 6,207 schools that were randomly assigned to APM or a control group at the beginning of 2016. Thus, there are schools outside the evaluation sample offering APM between 2016 and 2017. To distinguish these schools from the 182 APM schools in the evaluation sample, we classify *all* schools offering APM as "exposed". Exposed schools, therefore, include (i) The 182 APM schools in the evaluation sample; (ii) The 3,613 (3,795 – 182) APM schools that are not in the evaluation sample but were part of the larger randomized expansion; and (iii) Rural multi-grade schools that were not part of the randomized expansion but implemented APM before 2016.

**Table 3: Distribution of Observed Sample 1 Teachers by Their Destination School**

| 2016 School / Destination in 2017 | Treated | | Control | |
|---|---|---|---|---|
| | Number | % | Number | % |
| Same School | 179 | 0.818 | 200 | 0.848 |
| Exposed to APM | 13 | 0.059 | 13 | 0.055 |
| Treated school in the evaluation sample | 1 | 0.005 | 2 | 0.008 |
| Treated school out of the evaluation sample | 8 | 0.036 | 9 | 0.039 |
| Not randomized school | 4 | 0.018 | 2 | 0.008 |
| Not exposed to APM | 27 | 0.123 | 23 | 0.097 |
| Control school in the evaluation sample | 1 | 0.005 | 1 | 0.004 |
| Control school out of the evaluation sample | 4 | 0.018 | 4 | 0.017 |
| Not randomized school | 22 | 0.100 | 18 | 0.076 |
| Total | **219** | **1.00** | **236** | **1.00** |

**Table 4: Distribution of Sample 2 Teachers by Their School of Origin**

| 2017 School / Origin in 2016 | Treated | | Control | |
|---|---|---|---|---|
| | Number | % | Number | % |
| Same school | 179 | 0.599 | 200 | 0.587 |
| Exposed to APM | 33 | 0.110 | 34 | 0.100 |
| Treated school in the evaluation sample | 4 | 0.013 | 1 | 0.004 |
| Treated school out of the evaluation sample | 24 | 0.080 | 23 | 0.067 |
| Not randomized school | 5 | 0.017 | 10 | 0.029 |
| Not exposed to APM | 60 | 0.201 | 75 | 0.219 |
| Control school in the evaluation sample | 2 | 0.007 | 10 | 0.029 |
| Control school out of the evaluation sample | 10 | 0.033 | 13 | 0.038 |
| Not randomized school | 48 | 0.161 | 52 | 0.152 |
| Others [1/] | 27 | 0.090 | 32 | 0.094 |
| Total | **299** | **1.00** | **341** | **1.00** |

1/ Others: teachers whose school of origin cannot be identified due to lack of information.

Similarly, we classify schools that do not offer APM as "not exposed". These include: (i) The 182 non-APM schools in the evaluation sample; (ii) The 2,230 (2,432 – 182) non-APM schools that are not in the evaluation sample but were part of the larger randomized expansion; and (iii) Rural multi-grade schools that were not part of the randomization exercise and have never implemented APM. We use these classifications when describing the destination schools of Sample 1 teachers (Table 3) and the schools of origin of Sample 2 teachers (Table 4).

The distribution of teachers across the destination or origin schools is fairly well balanced between the control and treatment arms in both samples. Note that the percentage of Sample 1 teachers who stayed in their same school between year 1 (2016) and year 2 (2017), 82-85%, is much larger than the percentage of Sample 2 teachers who stayed in the same school in both years: 59-60%. This is likely due to the high rate of attrition among Sample 1 teachers (see Table 2), which was mainly due to difficulties in observing the teachers who changed schools between years 1 and 2. As a result, the proportion of teachers who stayed in their same school is over-represented in Sample 1. The information in Tables 3 and 4 can be used to estimate the proportion of each type of teacher (liker, disliker, mover or remainer) in the data. The details of these calculations are shown in Online Appendix 2. Table 5 presents the results.

The calculations in Table 5 for the proportions of each type of teacher in Sample 1 and Sample 2 can be viewed as particular realizations in the two samples we observe.

**Table 5: Estimated Proportions of the Four Types of Teachers in Samples 1 and 2**

| *Type of Teacher* | *Sample 1* | *Sample 2* |
|---|---|---|
| Remainer | 0.822 | 0.697 |
| Liker | -0.004 | -0.039 |
| Disliker | 0.026 | -0.023 |
| Mover | 0.157 | 0.365 |

To account for sampling variability, we drew 2,000 bootstrap samples (replications) from these two samples. Figure 3 shows the empirical distributions and mean values of the proportions of each type of teacher in Sample 1 (panel A) and Sample 2 (panel B) calculated in the same was as done for Table 5. There is no evidence of the presence of likers and dislikers; 90% confidence intervals for both of these types include zero, so the slightly negative estimates in Table 5 can be interpreted as estimates that are not significantly different from zero. We also conclude that the proportion of remainers and movers in both samples – $p^{R1} = 0.841$ and $p^{M1} = 0.159$ in Sample 1 and $p^{R2} = 0.639$ and $p^{M2} = 0.361$ in Sample 2[17] – are precisely estimated and are very far from zero.

It is possible to assess further the hypothesis of no likers or dislikers by checking for evidence of a composition effect. Figure 4 shows the correlation between the observed characteristics of Sample 2 teachers and the treatment status of the schools where they worked in year 2. If there is a sizeable composition effect, one is likely to find significant correlation between observed teacher characteristics and schools' treatment status. We find no evidence of such correlation. This is consistent with the absence of a composition effect, and thus with the absence of likers and dislikers.

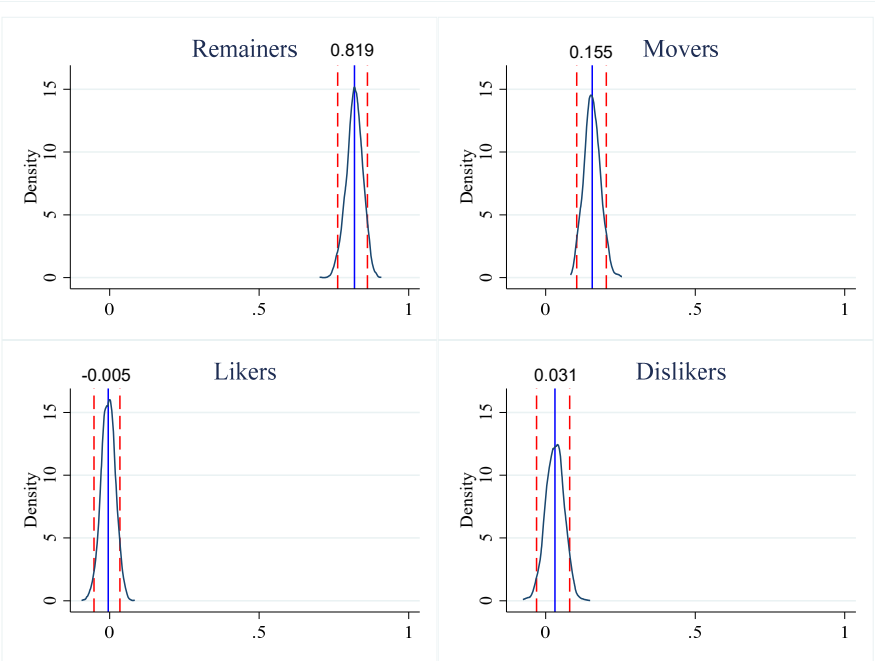
### 4.3 Ordinary Least Squares and Instrumental Variable Estimates

This subsection presents estimates of $E[y_{i2}|T_{i1} = 1] - E[y_{i2}|T_{i1} = 0]$, that is estimates of $\beta_1$ in equation (17), and $E[y_{i2}|T_{i2} = 1] - E[y_{i2}|T_{i2} = 0]$, that is estimates of $\beta_2$ in equation (21), using OLS regressions for Sample 1 and Sample 2 teachers, respectively. We also present the estimates obtained by regressing $y_i^2$ on the predicted years of treatment, instrumented by random assignment in year 1. As explained in subsection 3.4.1, this IV approach provides a consistent estimate of the $LATE_2$ parameter. For all

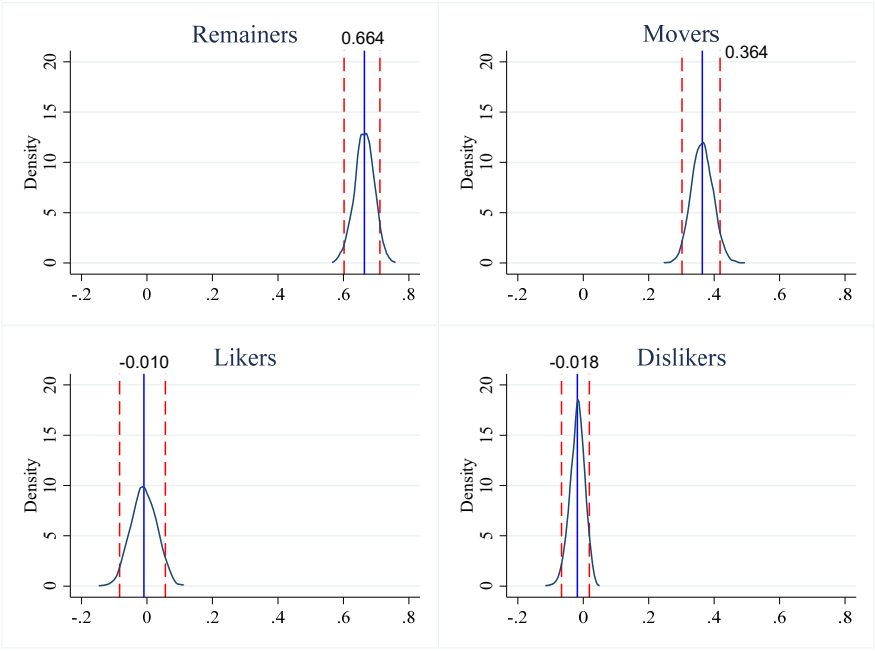
[17] These correspond to the mean values reported in Figure 3, after rescaling so that they sum to 1.

**Figure 3. Empirical Distributions of Proportions after 2,000 Replications of Sample 1 and Sample 2**

**Panel A. Proportions in Sample 1**

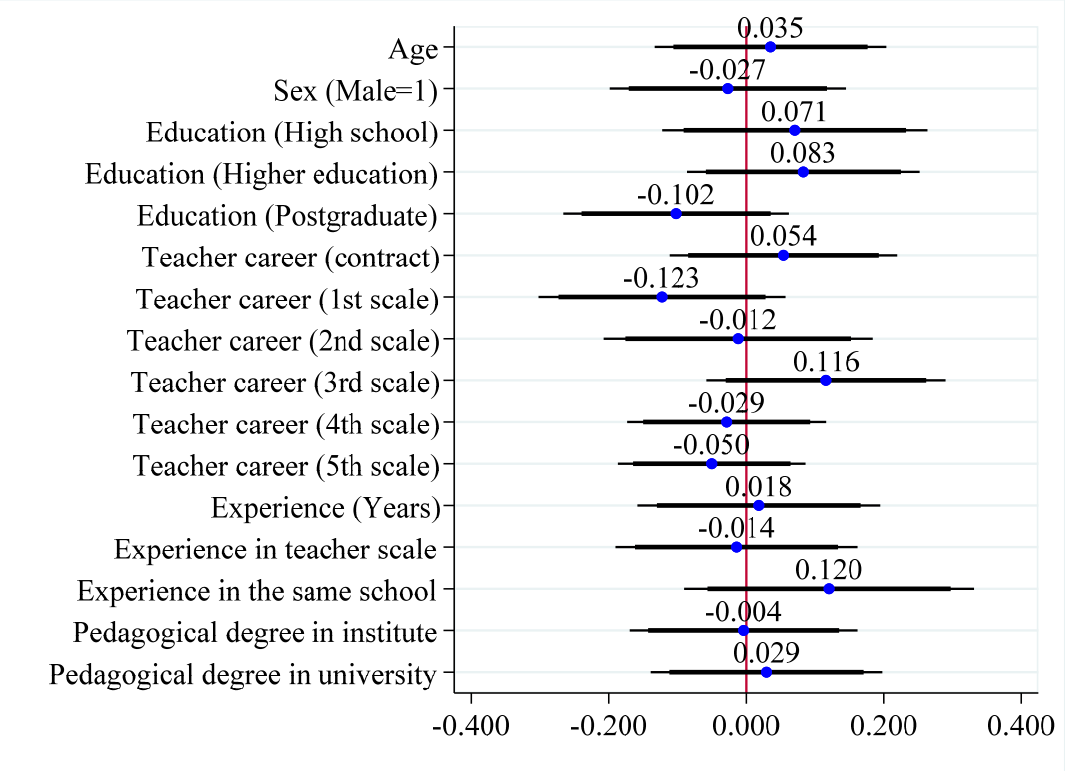



**Panel B. Proportions in Sample 2**




Note: Blue lines indicate the mean of the empirical distribution. Red lines indicate the 5th and 95th percentiles of the empirical distribution.

**Figure 4: Treatment Effects on the Composition of Teacher Characteristics among the Teachers Who Worked in Evaluation Sample Schools in 2017 (Sample 2)**




All regressions include UGEL fixed effects. Standard errors clustered at the school level.

Estimates indicate differences in the standardized characteristics of control and treatment groups. Thick and thin lines indicate 90% and 95% confidence intervals, respectively.

estimates, the dependent variable, $y_i^2$, is an index of pedagogical practices that averages the standardized scores of the eight indicators obtained from the classroom observations, as described in Section 2. We present estimates with and without teacher characteristics as covariates when using Sample 1.[18] Table 6 presents these results.

Before discussing the results, recall the analysis in subsection 4.2; it allows us to conclude that our population of teachers has no likers or dislikers. Note also that Section 3 shows that when likers and dislikers are absent then both $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ estimate ITT. Thus, all OLS estimates in Table 6 consistently estimate the same parameter: ITT.

[18] The use of teacher characteristics as covariates is appropriate only for Sample 1 because characteristics of Sample 2 teachers can be affected by the treatment. In Table A3.3 in the Appendix, we test for interactions between the treatment status and the characteristics of Sample 1 teachers. We find no evidence of heterogeneity by teacher experience, type of contract, position in the teacher career or sex. These results are important as they support the linearity assumption for the production function in equation (2).

**Table 6: Aggregate Skill: Ordinary Least Squares (OLS) estimates and IV Estimates**

| | Ordinary Least Squares Estimates | | | IV Estimates | |
|---|---|---|---|---|---|
| | Sample 1 | | Sample 2 | Sample 1 | |
| | (1) | (2) | (3) | (4) | (5) |
| **Treatment** | **0.287*** (0.108)** | **0.314*** (0.102)** | **0.195** (0.097)** | **0.159*** (0.054)** | **0.174*** (0.050)** |
| Experience | -- | 0.000 (0.009) | -- | -- | -0.000 (0.008) |
| Contract teacher | -- | 0.152 (0.162) | -- | -- | 0.145 (0.145) |
| Teacher career level | -- | 0.114** (0.046) | -- | -- | 0.113*** (0.041) |
| Sex (men = 1) | -- | -0.313*** (0.099) | -- | -- | -0.315*** (0.089) |
| Age | -- | -0.029*** (0.009) | -- | -- | -0.028*** (0.008) |
| $R^2$ | 0.29 | 0.37 | 0.23 | 0.29 | 0.37 |
| Sample Size | 455 | 455 | 640 | 455 | 455 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

All regressions include UGEL fixed effects. Standard errors clustered at the school level are in parentheses.

The first and second columns of Table 6 present estimates of $\hat{\beta}_{1,OLS}$. The estimates in Column (1), which do not control for teacher characteristics, indicate that offering APM for two years increases teachers' pedagogical skills by 0.29 standard deviations (s.d.). The estimate in Column (2), when teacher characteristics are added as covariates, is very similar: 0.31 s.d.. The estimate for $\hat{\beta}_{2,OLS}$ in Column (3), 0.20 s.d., is somewhat lower, even though $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ should both estimate ITT. Recall that remainers are very likely overrepresented in Sample 1, due to attrition. In contrast, the proportions of remainers and movers in Sample 2 should correspond to their proportions in the population of teachers in the 6,207 randomized expansion schools. Thus, $\hat{\beta}_{2,OLS}$ is our preferred estimate of ITT, the effect of *assigning* teachers to APM for two years on their aggregate pedagogical skill is 0.20 s.d.

In principle, if our Sample 1 and Sample 2 estimates differ because the proportions of movers and remainers in those two samples differ, we can use the different proportions of these two types of teachers shown in Table 5 to solve two equations with two unknowns: $\delta^R$ and $\delta^M$. This would allow us to estimate the average

treatment effects for both groups of teachers. However, our estimates of $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ are not significantly different from each other, so our data do not permit us to obtain precise estimates of $\delta^R$ and $\delta^M$.[19]

Our estimate that ITT = 0.20 sheds some light on other parameters of interest. Recall that in the absence of likers and dislikers $ATE_{teachers} \geq ATT_2 \geq ITT$. This implies that both the two-year effect of APM on the aggregate pedagogical practice of the average teacher ($ATE_{teachers}$) and the aggregate pedagogical practice of the teachers who were treated in year 2 ($ATT_2$) are at least as large as 0.2 s.d. In addition, recall that $ATE_{schools} = ATT_2$ when there are no likes and dislikers. This implies that the effect of the program after two years on the staff of the average school is also at least as large as 0.2 s.d.. This is consistent with the intuition presented at the end of subsection 3.3.3.

Columns (4) and (5) in Table 6 present our IV estimates of $LATE_2$ using sample 1. They show that one year of training increases by 0,16 to 0.17 standard deviations the pedagogical skill of compliers, when one considers all teachers to be compliers in year 1, but remainers receive a "double weight" because they were also compliers in year 2. Consistent with the fact that $LATE_2$ equals $ITT/(1 + p^R)$, this IV estimate is somewhat larger than (half of) the Sample 1 estimate of ITT in column (2).

### 4.4 The Effect of APM on Specific Pedagogical Practices

The discussion thus far has focused on the aggregate index of pedagogical skills, but one can also estimate the ITT of the program for each of the eight specific types of pedagogical skills shown in Table 1. Table 7 shows these results. To minimize spurious statistical significance that could arise from multiple hypothesis testing, Table 7 also presents adjusted p-values, using the stepdown method of Romano and Wolf (2016) to account for multiple hypothesis testing; these are in brackets below the standard errors.

The estimates in Table 7 indicate that the biggest impact of assigning teachers to the APM program, measured both by the size and the statistical significance of the estimated parameters, is on teachers' lesson planning; the point estimates are 0.34 for

[19] The difference between $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ is 0.119. We drew 2,000 bootstrap samples and calculated in each replication $\hat{\beta}_{1,OLS}$, $\hat{\beta}_{2,OLS}$ and their difference. We obtained that the bootstrapped standard error for the difference of betas is 0.083 and the bootstrapped p-value for the null hypothesis $\beta_1 - \beta_2 = 0$ is 0.224. Therefore, we cannot reject that $\beta_{1,OLS} = \beta_{2,OLS}$. In addition, in each replication, we calculated $\delta^R$ and $\delta^M$. We obtained that the bootstrapped $\delta^R$ is 0.201 and the bootstrapped $\delta^M$ is -0.181, and the standard errors are 0.078 and 0.335, respectively. Notice that $\delta^M$ is very imprecisely estimated.

**Table 7**
**Disaggregated Skills: Ordinary Least Squares Estimates**

| | (1) Lesson Planning | (2) Time Management | (3) Critical Thinking | (4) Student Participation | (5) Class Feedback | (6) Written Feedback | (7) Classroom Relationships | (8) Behavior Management |
|---|---|---|---|---|---|---|---|---|
| Panel A. Sample 1 | | | | | | | | |
| Treatment | 0.335 | 0.0811 | 0.268 | 0.168 | 0.193 | 0.138 | 0.0719 | 0.123 |
| | (0.105)*** | (0.105) | (0.099)*** | (0.107) | (0.104)* | (0.098) | (0.116) | (0.105) |
| | [0.018]** | [0.701] | [0.073]* | [0.488] | [0.348] | [0.511] | [0.701] | [0.580] |
| N | 448 | 450 | 450 | 450 | 450 | 448 | 450 | 450 |
| R-squared | 0.307 | 0.221 | 0.281 | 0.364 | 0.371 | 0.332 | 0.263 | 0.277 |
| Panel B. Sample 2 | | | | | | | | |
| Treatment | 0.375 | -0.0673 | 0.194 | 0.0627 | 0.0881 | 0.175 | 0.0225 | 0.0190 |
| | (0.088)*** | (0.092) | (0.094)** | (0.090) | (0.096) | (0.098)* | (0.095) | (0.089) |
| | [0.002]*** | [0.926] | [0.284] | [0.926] | [0.891] | [0.422] | [0.967] | [0.967] |
| N | 633 | 633 | 633 | 632 | 633 | 631 | 633 | 632 |
| R-squared | 0.245 | 0.171 | 0.200 | 0.260 | 0.277 | 0.236 | 0.209 | 0.238 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Note: Effects are measured in standard deviations. Regressions of Panel A include the following control variables: experience, contract teacher, teacher career level, sex and age. All regressions include UGEL fixed effects. Standard errors clustered at the school level are reported in parentheses and adjusted p-values for multiple hypotheses testing are reported in brackets. We calculate the adjusted p-values using the stepdown method of Romano and Wolf (2016).

Sample 1 and 0.38 for Sample 2. There is also evidence that APM increases teachers' pedagogical skills in developing their students' critical thinking, although the statistical significance is at best only marginal after controlling for multiple hypothesis testing.

# 5. The Effect of the APM Coaching on Student Learning

This section explores the impact of the APM coaching program on students' academic performance. Majerowicz and Montero (2021) evaluated the effect of APM on student learning and found significantly positive impacts on the 2016 National Student Evaluation (henceforth, ECE, its Spanish acronym). We complement these results by focusing on pedagogical practices as the relevant mechanism linking teacher coaching to student learning. Note that the program can raise students' learning by mechanisms other than their teachers' pedagogical practices. For example, the program may increase teachers' knowledge of the subjects they teach, or it can be seen as monitoring, which may motivate teachers to work harder and reduce their absenteeism.

We present two sets of results. First, we compare the test scores of students in the APM and non-APM schools in our evaluation sample that participated in the 2016 ECE.[20] This reveals whether the positive impacts found by Majerowicz and Montero (2021) also hold for the 364 schools we have used to estimate the effect of APM on teachers' pedagogical practices. Second, we assess whether the schools where teachers' pedagogical skills increased the most are also the schools with the largest increases in test scores. A positive correlation would indicate that changes in pedagogical practices are mediating at least part of the impact of teacher coaching on student learning.

About 40% of the 364 evaluation sample schools participated in the Grade 2 and 4 ECE assessments in 2016. This participation rate is low because the ECE is conducted only in schools with five or more students in a given grade and, by definition, multi-grade schools have relatively few students. This loss of about 60% of the schools may lead to lack of balance between the APM and non-APM schools with ECE scores. Of the 364 evaluation sample schools, 181 participated in either the Grade 2 or the Grade 4 ECE in 2016; 151 participated in the grade 2 ECE, 140 participated in the grade 4 ECE,

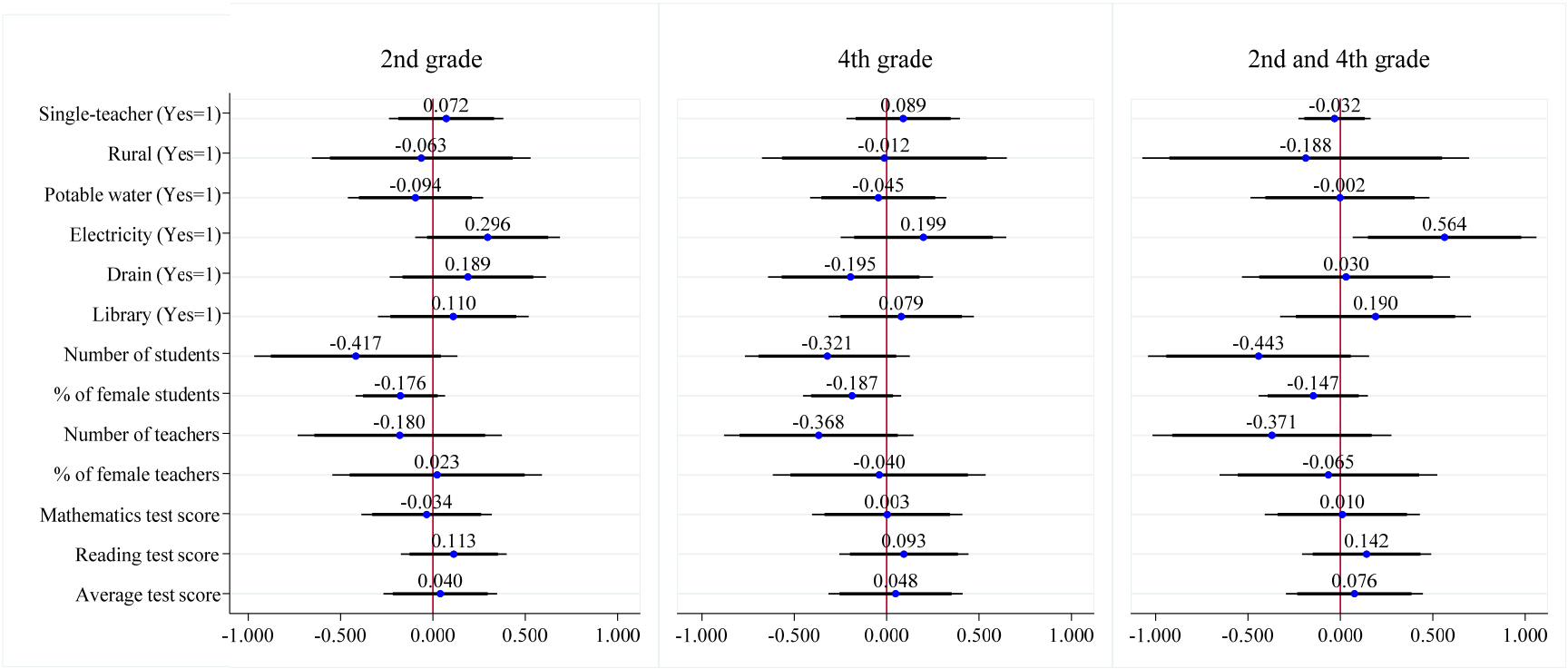
[20] National student evaluations assess student achievement in reading and mathematics using standardized tests given to students in Grades 2, 4 and 8. We focus on Grades 2 and 4 because APM coaching occurred in primary schools. The 2016 ECE covered Grades 2 and 4. Unfortunately, there was no ECE in 2017.

and 110 participated in both. Figure 5 shows the differences between the APM and non-APM schools in 10 basic school characteristics in the first half of 2016, for schools that participated in the grade 2 ECE, the grade 4 ECE and in both. It also includes the math and reading scores from the 2015 ECE for the 130 evaluation sample schools that participated in that assessment. We find no systematic difference between the APM and non-APM schools. Only one of the 33 differences is statistically significant, and only at the 5% level, which is about what one would expect from random chance. Overall, we conclude that the subsample of our 364 evaluation schools that has ECE data is unlikely to suffer from attrition bias.

Table 8 presents estimates of treatment effects of the APM coaching program on the average ECE scores. The ECE is taken near the end of the school year (which is also the end of the calendar year), so the 2016 ECE yields an estimate of the impact of the program after one year. All teachers complied with their random assignment in 2016, so this estimate can be interpreted as the average treatment effect (ATE) of one year of APM coaching on student learning. We also test for heterogeneity by the school's number of teachers because the program tended to focus on Grade 1 and 2 teachers for schools with more than one teacher. This means one can expect smaller effects for Grade 4 in schools with more teachers.

Columns (1) and (2) in Table 8 show results for the average 2016 ECE scores of Grade 2 students. There is a significant positive impact of 0.25 s.d. in Column (1), but no evidence of impact heterogeneity by the number of teachers in Column (2). Grade 4 results reveal no evidence of an average effect in Column (3), but there is heterogeneity by the number of teachers in Column (4). As expected, the effect on Grade 4 students weakens as the number of teachers rises, which increases the likelihood of being taught by a teacher who had less intense coaching (recall that APM focused on Grade 1 and Grade 2 teachers). The estimates in Column (4) show that, in schools with around two teachers, fourth graders' learning increased by an amount ($0.584 - 2 \times 0.154 = 0.276$) similar to the average effect on second graders (0.249). All these results are also found for mathematics and reading test scores separately (see Online Appendix 3, Tables A3.4 and A3.5).

**Figure 5. Balance in School Characteristics in the Sample of Schools that Participated in the National Student Evaluation in 2016 (2nd grade, 4th grade, and both grades)**




All regressions include UGEL fixed effects.

Estimates indicate differences in the standardized characteristics of control and treatment groups. Thick and thin lines indicate 90% and 95% confidence intervals, respectively.

Mathematics and Reading test scores refer to the average test scores obtained by the schools' 2nd grade students in the 2015 2nd grade National Student Evaluation. The average test score combines the results obtained in Mathematics and Reading.

**Table 8. Intention-to-Treat Estimates on Student Learning (Average Test Score Combining Mathematics and Reading)**

| | 2nd grade (2016) | | 4th grade (2016) | | 2nd and 4th grade (2016) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment (Yes=1) | 0.249** | 0.143 | 0.141 | 0.584** | 0.161* | 0.311** |
| | (0.126) | (0.200) | (0.150) | (0.257) | (0.0941) | (0.140) |
| Treatment×Number of teachers | | 0.075 | | -0.154** | | -0.028 |
| | | (0.066) | | (0.076) | | (0.047) |
| Number of teachers | | 0.086** | | 0.013 | | 0.058** |
| | | (0.039) | | (0.037) | | (0.028) |
| Constant | 0.435* | 0.138 | -0.235 | -0.284 | -0.570 | -0.785* |
| | (0.228) | (0.292) | (0.530) | (0.561) | (0.371) | (0.405) |
| Observations | 1,340 | 1,270 | 1,185 | 1,126 | 2,525 | 2,396 |
| Number of clusters | 151 | 138 | 140 | 129 | 181 | 161 |
| R-squared | 0.288 | 0.313 | 0.296 | 0.302 | 0.507 | 0.521 |
| UGEL FE | YES | YES | YES | YES | YES | YES |
| Grade FE | NO | NO | NO | NO | YES | YES |

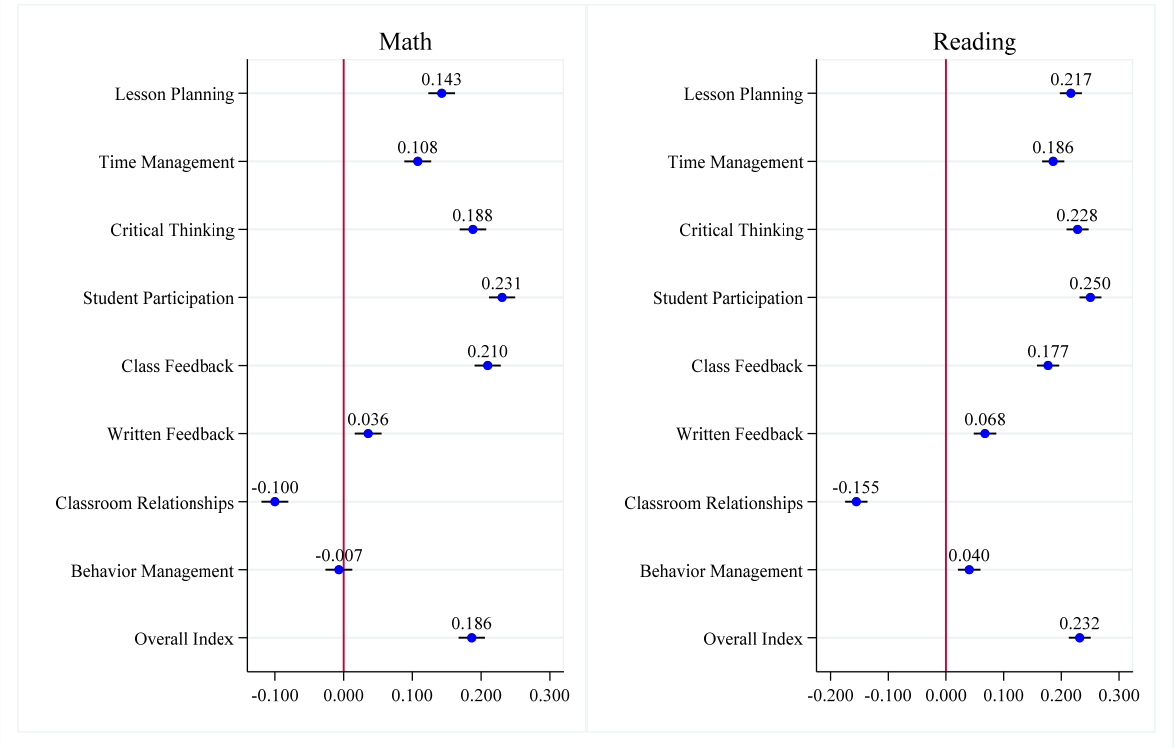
Estimates are in standard deviations (s.d.). Standard errors clustered at the school level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Did these increases in students ECE scores come about because APM coaching improved teachers' pedagogical practices? If so, the schools most affected in terms of teachers' pedagogical skills should also be the schools with the highest increases in test scores. We examine this by drawing bootstrap samples at the school level and checking whether the estimated impacts on teachers' pedagogical skills in the bootstrap samples are positively correlated with the estimates of the impacts on student performance on the ECE. This approach follows the methodology of Bennett, Naqvi and Schmidt (2018), who examined the possible correlation between the impact of microbe literacy on respondent's health and hygiene. If the effect on students' learning is mediated by the pedagogical practices, schools with teachers whose observed pedagogical skills benefited more from APM should also have the students that had the highest increase in test scores. That is, the difference in the composition of schools across the bootstrap samples should lead to a positive correlation between the estimated impacts of the program on pedagogical skills and on learning outcomes. Such correlations may also

indicate which pedagogical skills are most closely associated with increases in student learning. These correlations are shown in Figure 6.

Recall that these impacts on student learning are measured after one year of APM coaching (end of 2016) while the impacts on pedagogical skills are measured after two years (end of 2017). In addition, impacts on student learning are estimates of ATE while the impacts on pedagogical practices are ITT estimates due to teacher turnover between years 1 and 2 (2016 and 2017). Thus, this analysis relies on the plausible assumption that treatment effects on pedagogical practices that would have been observed at the end of 2016 are strongly positively correlated with the ITT effects estimated for 2017.

**Figure 6. Correlations of APM Impact on Teachers' Pedagogical Skills and Students' ECE Scores**




Note: Correlation coefficients and standard errors calculated based on the ITT estimates from10,000 bootstrap samples.

We find a positive and significant correlation between APM effects on student learning in math and reading and APM effects on the aggregate index of pedagogical practices (see last row of results in Figure 6). We also find that this positive correlation

is present for most of the eight more specific pedagogical skills. The skill with the strongest correlation for both mathematics and reading is encouraging student participation. Improvements in mathematics also have relatively strong correlations with higher skills in providing class feedback and encouraging students' critical thinking. For reading, we also find fairly strong correlation with critical thinking and lesson planning.

One unexpected result is the negative correlation between improving classroom relationships and the ECE test scores. This correlation is not particularly large, but it may indicate a tradeoff between encouraging good classroom relations and developing students' academic skills. It may also indicate that in schools with poorer student relations, trained teachers have a stronger response in terms of this pedagogical practice but are less capable of increasing student learning. Overall, the evidence in Figure 8 suggests that improving pedagogical practices is at least part of the mechanism linking the APM program to increased student learning.

# 6. Concluding Remarks

Teacher quality is a key determinant of student learning, but the quality of teachers is often low, especially in developing countries. Given the central role that human capital plays in economic growth and individuals' income and well-being, a key policy priority is to implement programs that increase teacher quality. The success of teacher training programs in raising teacher quality is, at best, mixed, but teacher coaching programs are a promising policy option.

We have estimated the effect of a large-scale teacher coaching program operating in a context of high teacher turnover in rural Peru on a broad range of pedagogical practices. This analysis contributes to the literature on teacher training and pedagogy by addressing the issues of scale and teacher turnover as potential threats to the effectiveness of coaching, and by presenting evidence that the general pedagogical skills of the current stock of teachers can be improved. We also explored the consequences of teacher turnover by developing an analytical framework that defines different types of treatment effects when teacher turnover is present and explains which treatment effects can be estimated.

Under the presence of turnover, the success of a teacher training or coaching program can be judged from two perspectives, the impact on the teachers who were

initially offered the treatment, regardless of whether they stay in their schools or move to a different school, and the impact on the teachers in treated schools after turnover has occurred. The first effect corresponds to the average intent to treat (ITT) and we show it can be estimated if one has a sample of teachers that follows them when they change schools, or using the data of teachers in treated and control schools after turnover has occurred *if* turnover is unrelated to the program. The second impact, which we call $ATE_{schools}$, cannot be estimated without bias even when turnover is unrelated to the program. However, we show that it is at least as large as the ITT for the teachers who were initially offered the treatment. We believe that this framework can be useful for future education evaluations carried out in contexts of high teacher turnover or, more generally, where treatments are offered at a cluster-level and service providers can change clusters while the intervention is still in progress.

We find that, after two years, the program has an (average) intent to treat (ITT) effect that increases teachers' pedagogical skills by 0.20 s.d. This effect is concentrated on two dimensions of the pedagogical practice: lesson planning and, to a lesser extent, encouraging students' critical thinking. We also estimated the effect of the program on student learning and found a positive effect after one year of coaching and a positive correlation between the size of this effect and the size of the effect on pedagogical skills. This is consistent with pedagogical skill being at least part of the mechanism linking teacher coaching to student learning.

This research also contributes to the discussion about which is the most cost-effective way to improve the pedagogical skill of teachers serving rural schools. Rural schools are often located in hard-to-reach areas that tend to be avoided by teachers if given a choice. One potential way to improve pedagogical skills and student learning in rural schools is by offering incentives to attract more talented teachers. The rural bonus scheme in Peru pursues this objective by offering an approximate 30% salary increase to those teachers who take a placement in a rural school. This bonus has had a small positive effect on the probability of filling a teacher vacancy but has shown no effects on learning outcomes (Castro and Esposito, 2021).

The cost of the coaching program evaluated in this study is around US$ 3,000 per teacher, per year. This is about 30% of the average annual salary of a primary school teacher in Peru, and it is similar to the wage premium offered by the bonus program, with two important differences: coaching is only a two-year investment (not a permanent salary increase), and it has proven effective to increase student learning.

Another potential way to improve pedagogical skills and student learning in rural schools is to offer incentives for (current) teachers to increase their productivity. The literature has shown that expensive policies based on large unconditional salary increases can reduce the number of teachers taking second jobs but have no effects on the productivity of teachers (de Ree et al., 2018). Pay-for-performance programs offer another alternative to improve teachers' productivity. The impact of these types of incentives has been examined in several low and middle-income countries, with mixed results. Very few studies, however, have estimated the effect of these programs in the context of a nation-wide intervention. A recent study by Bellés-Obrero and Lombardi (2019) evaluated the effect of a national pay-for-performance program implemented in 2015 in public secondary schools in Peru. The program, *Bono Escuela*, offers an additional monthly salary to the principal and teachers of the schools that rank in the top 20% of the national 8$^{th}$ grade student evaluation within their school district. The authors found no effect on student learning, as well as evidence that this lack of effect was related to teachers' uncertainty regarding which pedagogical practices lead to better scores.

Our results show that a large-scale coaching program can be an effective policy to improve the performance of existing teachers at a reasonable cost. Rather than offering incentives for teachers to devote more time and effort to the task (something that might not be effective if teachers lack the pedagogical skill), the results of this paper suggest that it is more effective to directly intervene to enhance their teaching skills.

# References


Albornoz, F., Anauati, M., Furman, M., Luzuriaga, M., Podesta, M., & Taylor, I. (2020). "Training to Teach Science: Experimental Evidence from Argentina." *The World Bank Economic Review,* 34(2), 393-417.

Angrist, J., G. Imbens & Rubin, D. (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434), 444-455.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., & Singh, N. (2021). "Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy, and Training." *American Economic Journal: Economic Policy*, 13(1), 36-66.

Bellés-Obrero, C., & Lombardi, M. (2019). "Teacher Performance Pay and Student Learning: Evidence from a Nationwide Program in Peru." *IZA Discussion Paper No. 12600*.

Bennett, D., Naqvi, A., & Schmidt, W. P. (2018). "Learning, Hygiene and Traditional Medicine." *The Economic Journal,* 128(612), F545-F574.

Bruns, B., Costa, L., & Cunha, N. (2018). Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil? *Economics of Education Review*, 64(1), 214-250.

Castro, J., & Esposito, B. (2021). "The Effect of Bonuses on Teacher Retention and Student Learning in Rural Schools: A Story of Spillovers." *Education, Finance and Policy*. Forthcoming.

Chetty, R., Friedman, J., & Rockoff, J. (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review,* 104(9), 2593-2632.

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2020). "How to Improve Teaching Practice? An Experimental Comparison of Centralized Training and In-Classroom Coaching." *Journal of Human Resources*, 55(3):926-962.

Clare, L., Garnier, H., Junker, B., & Correnti, R. (2010). "Investigating the Effectiveness of a Comprehensive Literacy Coaching Program in Schools with High Teacher Mobility." *The Elementary School Journal,* 111(1), 35-62.

Clotfelter, C., Ladd, H., & Vigdor, J. (2010). "Teacher Credentials and Student Achievement in High School: A Cross Subject Analysis with Fixed Effects." *Journal of Human Resources,* 45(3), 655-681.

Das, J., Dercon, S., Habyarimana, J., & Krishnan, P. (2007). “Teacher Shocks and Student Learning. Evidence from Zambia.” *Journal of Human Resources,* 42(4), 820-862.

de Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2018). “Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia.” *Quarterly Journal of Economics,* 133(2), 993-1039.

Evans, D., & Popova, A. (2016). “What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews.” *The World Bank Research Observer,* 31(2), 242-270.

Georgiadis, A., & Pitelis, C. N. (2016). “The Impact of Employees' and Managers' Training on the Performance of Small-and Medium-Sized Enterprises: Evidence from a Randomized Natural Experiment in the UK Service Sector.” *British Journal of Industrial Relations,* 54(2), 409-421.

Jukes, M., Turner, E., Dubeck, M., Halliday, K., Inyega, H., Wolf, S., ... and Brooker, S. (2017). “Improving literacy instruction in Kenya through teacher professional development and text messages support: A cluster randomized trial”. *Journal of Research on Educational Effectiveness*, 10(3):449-481.

Kotze, J., Fleisch, B., & Taylor, S. (2019). “Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa.” *International Journal of Educational Development*, 66, 203-213.

Kovner, C. T., Brewer, C. S., Fatehi, F., & Jun, J. (2014). “What Does Nurse Turnover Rate Mean and What is the Rate?” *Policy, Politics, & Nursing Practice*, 15(3-4), 64-71.

Kraft, M., Blazar, D., & Hogan, D. (2018). “The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence.” *Review of Educational Research,* 88(4), 547-588.

Loyalka, Prashant, Anna Popova, Guirong Li, Chengfang Liu, and Henry Shi (2019). “Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program.” *American Economic Journal: Applied Economics,* 11(3):128-154.

Lucas, Adrienne, Patrick McEwan, Moses Ngware and Moses Oketch. 2014. “Improving Early-grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda”. *Journal of Policy Analysis and Management* 33(4): 950-976.

Majerowicz, S., & Montero, R. (2021). “Can Teaching be Taught? Experimental Evidence from a Teacher Coaching Program in Peru.” *Working Paper*.

Popova, A., Evans, D., & Arancibia, V. (2016). “Training Teachers on the Job: What Works and How to Measure It.” Policy Research Working Paper 7834. The World Bank: Washington, DC.

Romano, J. P., and Wolf M. (2016). “Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing.” *Statistics & Probability Letters,* 113, 38-40.

Schaffner, Julie, Paul Glewwe and Uttam Sharma (2021). “Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed Methods Evaluation of an Unsuccessful Teacher Training Program.” Tufts University and University of Minnesota.

World Bank. (2018). *World Development Report: Learning to Realize Education's Promise.* The World Bank: Washington DC.

Zeitlin, Andrew (2021). “Teacher turnover in Rwanda.” *Journal of African Economies*, 30:1, 81-102.

# Online Appendix 1. Derivations for Subsections 3.3 and 3.4.

### Derivation of $ATT_1$

$ATT_1$ is defined in subsection 3.3 as $E[y^2_{T^2=1} - y^2_{T^2=0} | T^2 = 1]$. It can be expressed as a weighted average over teachers with $T^2 = 1$ who were randomly assigned to APM schools in year 1 and teachers with $T^2 = 1$ who were randomly assigned to non-APM schools in year 1:

$$ATT_1 = E[y^2_{T^2=1} - y^2_{T^2=0} | T^2 = 1] \qquad \text{(A1.1)}$$

$$= \frac{E[y^2_{1,1} - y^2_{1,0} | T^2 = 1, R^1 = 1]Pr[T^2 = 1 | R^1 = 1]Pr[R^1 = 1] + E[y^2_{0,1} - y^2_{0,0} | T^2 = 1, R^1 = 0]Pr[T^2 = 1 | R^1 = 0]Pr[R^1 = 0]}{Pr[T^2 = 1 | R^1 = 1]Pr[R^1 = 1] + Pr[T^2 = 1 | R^1 = 0]Pr[R^1 = 0]}$$

where $R^1$ indicates random assignment to an APM ($R^1 = 1$) or non-APM ($R^1 = 0$) school in year 1.

The terms Prob[$R^1 = 1$] and Prob[$R^1 = 0$] are important because we assume that there are a fixed number of teaching positions in APM (and non-APM) schools, which along with the proportion of schools that are APM schools, determine the proportion of likers and movers in the APM schools, and the proportion of dislikers and movers in the non-APM schools. Recall that $\tau$ is the proportion of schools, and thus the proportion of teachers, that were randomly assigned to the APM program in year 1. Thus Prob[$R^1 = 1$] = $\tau$ and Prob[$R^1 = 0$] = $1 - \tau$.

Next, consider the expressions for Prob[$T^2 = 1 | R^1 = 1$] and Prob[$T^2 = 1 | R^1 = 0$]. Define $p^R$, $p^L$, $p^D$ and $p^M$ as the proportions of teachers who are remainers, likers, dislikers and movers, respectively. Assume that likers get priority for APM schools, relative to movers, and the same holds for dislikers and non-APM schools, which implies that movers get whatever positions are "left over" after remainers, likers and dislikers have made their decisions. This is plausible since likers and dislikers have preferences for APM and non-APM schools, but movers are indifferent between the two types of schools. For APM schools, the proportion of positions in any given APM school that are available to movers is $1 - p^R - p^L/\tau$. Similarly, for non-APM schools the proportion of teaching positions that are available for movers is $1 - p^R - p^D/(1-\tau)$.[21] Since the proportion of APM schools is $\tau$, the proportion of teaching positions in APM schools that are available to movers is $(1 - p^R - p^L/\tau)\tau$. Similarly, since the proportion of non-APM schools is $1-\tau$, the proportion of teaching positions in non-APM schools that are available to movers is $(1 - p^R - p^D/(1-\tau))(1-\tau)$. Given that the four proportions of the different types of teachers must sum to 1, is it straightforward to show that these two proportion probabilities sum to $p^M$.

Consider movers in APM schools ($R^1 = 1$) in year 1. The proportion of teachers in APM schools (and non-APM schools) in year 1 who are movers is $p^M$. The expressions above

[21] Technically, we assume that the proportions of the four types of teachers do not change from year 1 to year 2 in the in the 6,207 multi-grade schools included in the "randomized expansion" that occurred in 2016 See the text for further discussion of this assumption and for indirect evidence supporting it.

imply that the probability that a mover in an APM (or non-APM) school ends up in an APM school in year 2 is $[(1 - p^R - p^L/\tau)\tau]/[(1 - p^R - p^L/\tau)\tau + (1 - p^R - p^D/(1-\tau))(1-\tau)] = (1 - p^R - p^L/\tau)\tau/p^M$. Thus we have:

$$\text{Prob}[T^2 = 1 | R = 1] = p^R + p^L + p^M \times \text{Prob}[\text{Mover stays in an APM school}] \quad \text{(A1.2)}$$

$$= p^R + p^L + (1 - p^R - p^L/\tau)\tau$$

$$\text{Prob}[T^2 = 1 | R = 0] = p^L + (1 - p^R - p^L/\tau)\tau \quad \text{(A1.3)}$$

The calculation of the different potential outcomes for y can then be expressed as:

$$E[y^2_{1,1} | T^2 = 1, R^1 = 1] \quad \text{(A1.4)}$$

$$= [(\theta^{2,R} + 2\delta^R)p^R + (\theta^{2,L} + 2\delta^L)p^L + (\theta^{2,M} + 2\delta^M)(1 - p^R - p^L/\tau)\tau]/(p^R + p^L + (1 - p^R - p^L/\tau)\tau)$$

$$E[y^2_{1,0} | T^2 = 1, R^1 = 1] \quad \text{(A1.5)}$$

$$= [(\theta^{2,R} + \delta^R)p^R + (\theta^{2,L} + \delta^L)p^L + (\theta^{2,M} + \delta^M)(1 - p^R - p^L/\tau)\tau]/(p^R + p^L + (1 - p^R - p^L/\tau)\tau)$$

$$E[y^2_{0,1} | T^2 = 1, R^1 = 0] \quad \text{(A1.6)}$$

$$= [(\theta^{2,L} + \delta^L)p^L + (\theta^{2,M} + \delta^M)(1 - p^R - p^L/\tau)\tau]/(p^L + (1 - p^R - p^L/\tau)\tau)$$

$$E[y^2_{0,0} | T^2 = 1, R^1 = 0] \quad \text{(A1.7)}$$

$$= [\theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau]/(p^L + (1 - p^R - p^L/\tau)\tau)$$

Putting this all together gives:

$$ATT_1 = E[y^2_{T^2=1} - y^2_{T^2=0} | T^2 = 1] \quad \text{(A1.8)}$$

$$= \frac{E[y^2_{1,1} - y^2_{1,0} | T^2 = 1, R^1 = 1]\Pr[T^2 = 1 | R^1 = 1]\Pr[R^1 = 1] + E[y^2_{0,1} - y^2_{0,0} | T^2 = 1, R^1 = 0]\Pr[T^2 = 1 | R^1 = 0]\Pr[R^1 = 0]}{\Pr[T^2 = 1 | R^1 = 1]\Pr[R^1 = 1] + \Pr[T^2 = 1 | R^1 = 0]\Pr[R^1 = 0]}$$

$$= \{[\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau)]\tau + [\delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau](1-\tau)\}/(\tau p^R + p^L + (1 - p^R - p^L/\tau)\tau)$$

$$= [\tau\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau)]/\tau$$

$$= \delta^R p^R + \delta^L(p^L/\tau) + \delta^M(1 - p^R - p^L/\tau)$$

**Derivation of $ATT_2$**

$ATT_2$ is defined in subsection 3.3 as $E[y^2_{T^2=1} - y^2_{0,0} | T^2 = 1]$. As with $ATT_1$, it can be expressed as a weighted average over teachers with $T^2 = 1$ who were randomly assigned to APM schools in year 1 and teachers with $T^2 = 1$ who were randomly assigned to non-APM schools in year 1:

$$ATT_2 = E[y^2_{T^2=1} - y^2_{0,0} | T^2 = 1] \quad \text{(A1.9)}$$

$$= \frac{E\left[y_{1,1}^2 - y_{0,0}^2 \middle| T^2 = 1, R^1 = 1\right]Pr[T^2 = 1|\, R^1 = 1]Pr[R^1 = 1] + E\left[y_{0,1}^2 - y_{0,0}^2 \middle| T^2 = 1, R^1 = 0\right]Pr[T^2 = 1|\, R^1 = 0]Pr[R^1 = 0]}{Pr[T^2 = 1|\, R^1 = 1]Pr[R^1 = 1] + Pr[T^2 = 1|\, R^1 = 0]Pr[R^1 = 0]}$$

The only term in $ATT_2$ that was not shown above as a component for $ATT_1$ is:

$$E[y_{0,0}^2|\, T^2 = 1, R^1 = 1] \qquad \text{(A1.10)}$$

$$= [\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau]/(p^R + p^L + (1 - p^R - p^L/\tau)\tau)$$

Putting this all together yields:

$$ATT_2 = E[y_{T^2=1}^2 - y_{0,0}^2|\, T^2 = 1] \qquad \text{(A1.11)}$$

$$= \{[2\delta^R p^R + 2\delta^L p^L + 2\delta^M(1 - p^R - p^L/\tau)\tau] + [\delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau](1-\tau)\}/(\tau p^R + p^L + (1 - p^R - p^L/\tau)\tau)$$

$$= [2\delta^R p^R \tau + \delta^L p^L(1+\tau) + \delta^M(1 - p^R - p^L/\tau)\tau(1+\tau)]/\tau$$

$$= 2\delta^R p^R + (1+\tau)\delta^L(p^L/\tau) + (1+\tau)\delta^M(1 - p^R - p^L/\tau)$$

**Derivation of ITT**

ITT is defined in subsection 3.3 as $E[y^2|\, R^1 = 1] - E[y^2|\, R^1 = 0]$. Each of these two components can be expressed as a weighted average over teachers for whom $T^2 = 1$ and teachers for whom $T^2 = 0$:

$$ITT = E[y^2|\, R^1 = 1] - E[y^2|\, R^1 = 0] \qquad \text{(A1.12)}$$

$$= E[y_{1,1}^2|\, T^2 = 1, R^1 = 1] \times Pr[T^2 = 1|\, R^1 = 1] + E[y_{1,0}^2|\, T^2 = 0, R^1 = 1] \times Pr[T^2 = 0|\, R^1 = 1]$$

$$- \{E[y_{0,1}^2|\, T^2 = 1, R^1 = 0] \times Pr[T^2 = 1|\, R^1 = 0] + E[y_{0,0}^2|\, T^2 = 0, R^1 = 0] \times Pr[T^2 = 0|\, R^1 = 0]\}$$

Using equation (4), $y_i^2 = \theta_i^2 + \delta_i(T_i^1 + T_i^2)$, and noting that $\tau$ is the fraction of APM schools:

$$Pr[T^2 = 1|\, R^1 = 1] = p^R + p^L + (1 - p^R - p^L/\tau)\tau = p^R + (1-p^R)\tau \qquad \text{(A1.13)}$$

$$E[y_{1,1}^2|\, T^2 = 1, R^1 = 1] \qquad \text{(A.14)}$$

$$= [\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + 2(\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau)]/(p^R + (1-p^R)\tau)$$

$$Pr[T^2 = 0|\, R^1 = 1] = p^D + (1 - p^R - p^D/(1-\tau))(1-\tau) = (1-\tau)(1 - p^R) \qquad \text{(A1.15)}$$

$$E[y_{1,0}^2|\, T^2 = 0, R^1 = 1] \qquad \text{(A.16)}$$

$$= [\theta^{2,D}p^D + \theta^{2,M}(1 - p^R - p^D/(1-\tau)) \times (1-\tau) + \delta^D p^D + \delta^M(1 - p^R - p^D/(1-\tau)) \times (1-\tau)]/(1-\tau)(1 - p^R)$$

$$Pr[T^2 = 1|\, R^1 = 0] = p^L + (1 - p^R - p^L/\tau)\tau = (1 - p^R)\tau \qquad \text{(A1.17)}$$

$$E[y^2_{0,1}|\ T^2 = 1, R^1 = 0] \qquad \text{(A1.18)}$$

$$= [\theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau]/(1 - p^R)\tau$$

$$\Pr[T^2 = 0|\ R^1 = 0] = p^R + p^D + (1 - p^R - p^D/(1-\tau))\times(1-\tau) = 1 - \tau(1 - p^R) \qquad \text{(A.19)}$$

$$E[y^2_{0,0}|\ T^2 = 0, R^1 = 0] \qquad \text{(A1.20)}$$

$$= [\theta^{2,R}p^R + \theta^{2,D}p^D + \theta^{2,M}(1 - p^R - p^D/(1-\tau))\times(1-\tau)]/(1 - \tau(1 - p^R))$$

Inserting all of these into the equation above for ITT gives:

$$ITT = \theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + 2(\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau) \quad \text{(A1.21)}$$

$$+ \theta^{2,D}p^D + \theta^{2,M}(1 - p^R - p^D/(1-\tau))\times(1-\tau) + \delta^D p^D + \delta^M(1 - p^R - p^D/(1-\tau))\times(1-\tau)$$

$$- [\theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau + \theta^{2,R}p^R + \theta^{2,D}p^D + \theta^{2,M}(1 - p^R - p^D/(1-\tau))\times(1-\tau)]$$

$$= 2(\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau) + \delta^D p^D + \delta^M(1 - p^R - p^D/(1-\tau))\times(1-\tau) - [\delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau]$$

$$= 2\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M[\tau(1 - p^R) - p^L] + \delta^M[(1 - p^R)(1-\tau) - p^D]$$

$$= 2\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M(1 - p^R - p^L - p^D)$$

$$= 2\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p_1^M$$

$$= \bar{\delta} + \delta^R p^R$$

**Derivation of $ATE_{schools}$**

$ATE_{schools}$ is defined in subsection 3.3 as $E[y^2_{T^2=1}|\ T^2 = 1] - E[y^2_{0,0}]$. The first term can be expressed as a weighted average over teachers with $T^2 = 1$ who were randomly assigned to APM schools in year 1 and teachers with $T^2 = 1$ who were randomly assigned to non-APM schools in year 1:

$$E[y^2_{T^2=1}|\ T^2 = 1] \qquad \text{(A1.22)}$$

$$= \frac{E[y^2_{1,1}|T^2 = 1, R^1 = 1] \times \Pr[T^2 = 1|\ R^1 = 1] \times \Pr[R^1 = 1] + E[y^2_{0,1}\ |T^2 = 1, R^1 = 0] \times \Pr[T^2 = 1|\ R^1 = 0] \times \Pr[R^1 = 0]}{\Pr[T^2 = 1|\ R^1 = 1] \times \Pr[R^1 = 1] + \Pr[T^2 = 1|\ R^1 = 0] \times \Pr[R^1 = 0]}$$

As in the derivation of $ATT_1$ (see equation (A1.8)), the denominator simply equals $\tau$. Noting that $\Pr[T^2 = 1|\ R^1 = 1] = p^R + (1-p^R)\tau$ and $\Pr[T^2 = 1|\ R^1 = 0] = (1 - p^R)\tau$, the numerator of (A1.22) is:

$$[\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + 2(\delta^R p^R + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau)]\times\tau \quad \text{(A1.23)}$$

$$+ [\theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + \delta^L p^L + \delta^M(1 - p^R - p^L/\tau)\tau]\times(1-\tau)$$

Subtracting this numerator by the denominator ($\tau$) and subtracting, $E[y^2_{0,0}]$, which equals $\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,D}p^D + \theta^{2,M}p^M$, yields:

$$ATE_{schools} \qquad (A1.24)$$

$$= [\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,M}(1 - p^R - p^L/\tau)\tau + 2(\delta^R p^R + \delta^L p^L + \delta^M(1 - p_1^R - p^L/\tau)\tau)]$$

$$+ [\theta^{2,L}p^L/\tau + \theta^{2,M}(1 - p^R - p^L/\tau) + \delta^L p^L/\tau + \delta^M(1 - p^R - p^L/\tau)]\times(1-\tau)$$

$$- [\theta^{2,R}p^R + \theta^{2,L}p^L + \theta^{2,D}p^D + \theta^{2,M}p^M]$$

$$= 2\delta^R p^R + \delta^L p^L[2 + (1-\tau)/\tau] + \delta^M[2\tau(1 - p^R - p^L/\tau) + (1-\tau)(1 - p^R - p^L/\tau)]$$

$$+ \theta^{2,L}p^L[1+ (1-\tau)/\tau - 1] - \theta^{2,D}p^D + \theta^{2,M}[(1 - p^R - p^L/\tau)(\tau + (1-\tau)) - p^M]$$

$$= 2\delta^R p^R + \delta^L p^L[(1+\tau)/\tau] + \delta^M(1+\tau)(1 - p^R - p^L/\tau)$$

$$+ \theta^{2,L}p^L(1-\tau)/\tau - \theta^{2,D}p^D + \theta^{2,M}(1 - p^R - p^L/\tau - p^M)$$

**Derivation of Regressing $y_i^2$ on $T_i^1$ (Equation (17)) for Sample 2 Teachers**

The regression equation is $y_i^2 = \alpha_1 + \beta_1 T_i^1 + \varepsilon_{1i}$, applied to Sample 2 teachers. An implicit assumption in this derivation is that the proportion terms ($p^R$, $p^L$, $p^D$ and $p^M$) for the Sample 2 teachers are the same in both year 1 and year 2, which implies that any movement of teachers into or out of the 6,207 randomized expansion schools is uncorrelated with the type of teacher.

This regression produces the following estimate for $\beta_1$:

$$\{E[y_i^2 | T_i^1 = 1, T_i^2 = 1]\times Prob[T_i^2 = 1 | T_i^1 = 1] + E[y_i^2 | T_i^1 = 1, T_i^2 = 0]\times Prob[T_i^2 = 0 | T_i^1 = 1]\}$$

$$- \{E[y_i^2 | T_i^1 = 0, T_i^2 = 1]\times Prob[T_i^2 = 1 | T_i^1 = 0] + E[y_i^2 | T_i^1 = 0, T_i^2 = 0]\times Prob[T_i^2 = 0 | T_i^1 = 0]\}$$

The terms in the above expression are the following functions of the $p$, $\delta$ and $\theta$ terms:

$Prob[T_i^2 = 1 | T_i^1 = 1] = p^R + p^L + \tau p^M$ (teachers in treated schools who stay in those schools)

$E[y_i^2 | T_i^1 = 1, T_i^2 = 1] = [(\theta^{2,R} + 2\delta^R)p^R + (\theta^{2,L} + 2\delta^L)p^L + (\theta^{2,M} + 2\delta^M)\tau p^M]/(p^R + p^L + \tau p^M)$

$Prob[T_i^2 = 0 | T_i^1 = 1] = p^D + (1 - \tau)p^M$ (teachers in treated schools who move to untreated schools)

$E[y_i^2 | T_i^1 = 1, T_i^2 = 0] = [(\theta^{2,D} + \delta^D)p^D + (\theta^{2,M} + \delta^M)(1 - \tau)p^M]/(p^D + (1 - \tau)p^M)$

$Prob[T_i^2 = 1 | T_i^1 = 0] = p^L + \tau p^M$ (teachers in untreated schools who move to a treated school)

$E[y_i^2 | T_i^1 = 0,\ T_i^2 = 1] = [(\theta^{2,L} + \delta^L)p^L + (\theta^{2,M} + \delta^M)\tau p^M]/(p^L + \tau p^M)$

Prob$[T_i^2 = 0 | T_i^1 = 0]$ = $\mathrm{p^R} + \mathrm{p^D} + (1-\tau)\mathrm{p^M}$ (teachers in untreated schools who stay in those schools)

$$\mathrm{E}[y_i^2 \big| T_i^1 = 0, T_i^2 = 0] = [\theta^{2,R}\mathrm{p}^R + \theta^{2,D}\mathrm{p^D} + \theta^{2,M}(1-\tau)\mathrm{p^M}]/(\mathrm{p^R} + \mathrm{p^D} + (1-\tau)\mathrm{p^M})$$

Inserting all of these expressions into the above estimate for $\beta_1$ yields (note that "Prob" terms are canceled out by the denominators of the conditional expectation terms):

$$\{(\theta^{2,R} + 2\delta^R)\mathrm{p^R} + (\theta^{2,L} + 2\delta^L)\mathrm{p^L} + (\theta^{2,M} + 2\delta^M)\tau\mathrm{p^M} + (\theta^{2,D} + \delta^D)\mathrm{p^D} + (\theta^{2,M} + \delta^M)(1-\tau)\mathrm{p^M}\}$$

$$- \{(\theta^{2,L} + \delta^L)\mathrm{p^L} + (\theta^{2,M} + \delta^M)\tau\mathrm{p^M} + \theta^{2,R}\mathrm{p}^R + \theta^{2,D}\mathrm{p^D} + \theta^{2,M}(1-\tau)\mathrm{p^M}\}$$

$$= \{(\theta^{2,R} + 2\delta^R)\mathrm{p^R} + (\theta^{2,L} + 2\delta^L)\mathrm{p^L} + (\theta^{2,D} + \delta^D)\mathrm{p^D} + (\theta^{2,M} + (1+\tau)\delta^M)\mathrm{p^M}\}$$

$$- \{(\theta^{2,L} + \delta^L)\mathrm{p^L} + \theta^{2,R}\mathrm{p}^R + \theta^{2,D}\mathrm{p^D} + (\theta^{2,M} + \tau\delta^M)\mathrm{p^M}\}$$

$$= 2\delta^R\mathrm{p^R} + 2\delta^L\mathrm{p^L} + \delta^D\mathrm{p^D} + (1+\tau)\delta^M\mathrm{p^M} - \delta^L\mathrm{p^L} - \tau\delta^M\mathrm{p^M}$$

$$= 2\delta^R\mathrm{p^R} + \delta^L\mathrm{p^L} + \delta^D\mathrm{p^D} + \delta^M\mathrm{p^M} -$$

$$= \overline{\delta_1} + \delta^R\mathrm{p^R} \text{ (which is ITT).}$$

# Online Appendix 2. Derivations for Subsection 4.2

**Derivations for Sample 1 teachers**

For Sample 1 teachers in our data, we denote the proportions of likers, dislikers, movers and remainers by $p^{L1}$, $p^{D1}$, $p^{M1}$ and $p^{R1}$, respectively. The "1" denotes Sample 1 teachers; these proportions could differ from the proportions in the 6,207 randomized expansion schools due to differential attrition among the four types of teachers.

In any given school, the number of teaching positions is fixed, so the total number of teachers does not change. Recall that, in general, movers are assumed to be indifferent between APM and non-APM schools, and so likers and dislikers "move first" in terms of the schools that they want to switch to, after which movers can have the remaining teaching positions in a given school. Among all schools that movers move to in year 2, let $\mu$ be the proportion that are treated schools. Note that this is equal to $(1 - p^{R1} - p^{L1}/\tau)*\tau/p^{M1}$; it may not equal $\tau$ because movers end up in the schools with unfilled positions, and the number of those positions in APM and non-APM will depend on the proportion of likers and dislikers in the population of teachers. It will be seen below that $\mu$ can be calculated from the data.

Consider teachers in Sample 1 schools. Table A2.1 shows how teachers move, or do not move, from year 1 to year 2 based on the type of teacher and random assignment in year 1, and how this is related to the proportion of the four types of teachers.

**Table A2.1: Move Decisions of Sample 1 Teachers**

| | Move decision (observed) | Likers | Movers | Remainers | Dislikers | Row Sum (observed) |
|---|---|---|---|---|---|---|
| Assigned to APM school | move to APM school | 0 | $p^{M1}\mu$ | 0 | 0 | a |
| | move to non-APM school | 0 | $p^{M1}(1-\mu)$ | 0 | $p^{D1}$ | b |
| | stayed in same school | $p^{L1}$ | 0 | $p^{R1}$ | 0 | c |
| Assigned to non-APM school | move to APM school | $p^{L1}$ | $p^{M1}\mu$ | 0 | 0 | d |
| | move to non-APM school | 0 | $p^{M1}(1-\mu)$ | 0 | 0 | e |
| | stayed in same school | 0 | 0 | $p^{R1}$ | $p^{D1}$ | f |

The first three rows of Table A2.1 show how teachers assigned to an APM school in year 1 move, or do not move, to a different school in year 2. For example, all likers in these schools are satisfied with being assigned to a treated school, and they are assumed to stay in that school (as opposed to moving to a different treated school), so of all teachers assigned to APM schools, $p^{L1}$ are likers, and all of them stay in the school to which they were assigned. The same is true for remainers, so of all teachers assigned to APM schools, $p^{R1}$ are remainers, and all of them stay in the school to which they were assigned. Analogously, of all teachers assigned to an APM school, $p^{D1}$ are dislikers, and all of them move to a non-APM. Finally, movers move randomly out of the school

to which they were initially assigned, and given the schools available to them $p^{M1}\mu$ end up in APM schools and $p^{M1}(1-\mu)$ end up in non-APM schools.

The identity of the four different types of teachers is not observed. Of all teachers in the APM schools in year 1, one does observe the proportions that moved to an APM school, moved to a non-APM school, or did not move at all. These proportions are shown as a, b and c in the last column of Table A2.1, and they sum to one. Thus the relationship between the observed proportions in Table A2.1 and the unobserved values for $p^{L1}$, $p^{M1}$, $p^{R1}$ and $p^{D1}$, as well as the unobserved value for $\mu$, is governed by the following three equations:

$$p^{M1}\mu = a \quad \text{(A2.1)}$$
$$p^{M1}(1-\mu) + p^{D1} = b \quad \text{(A2.2)}$$
$$p^{L1} + p^{R1} = c \quad \text{(A2.3)}$$

The same types of relationships hold for the Sample 1 teachers who were assigned to non-APM schools, and these are shown in the last three lines of Table A3.1. These three lines generate the following three equations:

$$p^{L1} + p^{M1}\mu = d \quad \text{(A2.4)}$$
$$p^{M1}(1-\mu) = e \quad \text{(A2.5)}$$
$$p^{R1} + p^{D1} = f \quad \text{(A2.6)}$$

This gives six equations and five unknowns ($\mu$, $p^{L1}$, $p^{M1}$, $p^{R1}$ and $p^{D1}$). In fact, these six equations are not independent, which can be shown in two steps.

Step 1. Add equations (A.2.1) and (A.2.3), and subtract equation (A.2.4):

$$p^{M1}\mu + p^{L1} + p^{R1} - p^{L1} - p^{M1}\mu = a + c - d \quad \text{(A2.7)}$$

$$p^{R1} = a + c - d \quad \text{(A2.8)}$$

Step 2. Add equation (A.2.2) and subtract equation (A.2.5):

$$p^{R1} + p^{M1}(1-\mu) + p^{D1} - p^{M1}(1-\mu) = a + b + c - d - e \quad \text{(A2.9)}$$

$$p^{R1} + p^{D1} = 1 - d - e \text{ (since } a + b + c = 1) \quad \text{(A2.10)}$$

$$p^{R1} + p^{D1} = f \text{ (since } d + e + f = 1), \text{ which is equation (A.2.6))} \quad \text{(A2.11)}$$

Thus there are five independent linear equations and five unknowns ($p^{L1}$, $p^{M1}\mu$, $p^{M1}(1-\mu)$, $p^{R1}$ and $p^{D1}$).[22] To solve for the p's and $\mu$, the proportions a, b, c, d, e and f can be calculated using Table 3 in the main text:

$$a = 13/219 = 0.0594 \quad \text{(A2.12)}$$

$$b = 27/219 = 0.1233 \quad \text{(A2.13)}$$

[22] In fact, there is another equation, which is that the four proportions of teachers sum to 1, and thus these five unknowns sum to 1. Yet this equation is redundant because a + b + c = 1 and d + e + f =1 both imply that these four proportions sum to 1.

$$c = 179/219 = 0.8174 \qquad \text{(A2.14)}$$

$$d = 13/236 = 0.0551 \qquad \text{(A2.15)}$$

$$e = 23/236 = 0.0975 \qquad \text{(A2.16)}$$

$$f = 200/236 = 0.8475 \qquad \text{(A2.17)}$$

Returning to the six equations:

$$p^{M1}\mu = a = 0.0594 \qquad \text{(A2.18)}$$

$$p^{M1}(1-\mu) = 0.0975 \qquad \text{(A2.19)}$$

So total movers ($p^{M1}$) equals:

$$0.0594 + 0.0975 = 0.1569 \qquad \text{(A2.20)}$$

The remaining unknowns are solved as follows:

$$p^{D1} = b - e = 0.1233 - 0.0975 = 0.0258 \qquad \text{(A2.21)}$$

$$p^{L1} = d - a = 0.0551 - 0.0594 = -0.0043 \qquad \text{(A2.22)}$$

$$p^{R1} = c - (d - a) = 0.8174 - (0.0551 - 0.0594) = 0.8174 - (-0.0043) = 0.8217 \qquad \text{(A2.23)}$$

$$\mu = a/p^{M1} = 0.0594/0.1569 = 0.3786 \qquad \text{(A2.24)}$$

**Derivations for Sample 2 Teachers**

The information in Table 4 in the main text for the Sample 2 teachers can used to estimate $p^{L2}$, $p^{M2}$, $p^{R2}$, $p^{D2}$ and $\mu$. These estimates, which include a "2" superscript, are likely to be different from those obtained using the information for Sample 1 teachers because the proportion of teachers who stay in their same school between years 1 and 2 are likely to be overrepresented in Sample 1. Recalling that $\tau$ is be the proportion of teaching positions in APM schools, and that the number of teaching positions in those schools is fixed, the proportion of likers in APM schools in year 2 will be $p^{L2}/\tau$. Similarly, the proportion of dislikers in non-APM schools in year 2 will be $p^{D2}/(1-\tau)$. Table A2.2 shows how Sample 2 teachers are allocated to APM and non-APM schools:

**Table A2.2: Move Decisions of Sample 2 Teachers**

| | School type in year 1 (observed) | Likers | Movers | Remainers | Dislikers | Row Sum (observed) |
|---|---|---|---|---|---|---|
| APM school in year 2 | Treated | 0 | $(1-p^{R2} - p^{L2}/\tau)\mu$ | 0 | 0 | "a" |
| | Control | $(p^{L2}/\tau)(1-\tau)$ | $(1-p^{R2} - p^{L2}/\tau)(1-\mu)$ | 0 | 0 | "b" |
| | Same | $(p^{L2}/\tau)\tau = p^{L2}$ | 0 | $p^{R2}$ | 0 | "c" |
| Non-APM school in year 2 | Treated | 0 | $(1-p^{R2} - p^{D2}/(1-\tau))\mu$ | 0 | $(p^{D2}/(1-\tau))\tau$ | "d" |
| | Control | 0 | $(1-p^{R2} - p^{D2}/(1-\tau))(1-\mu)$ | 0 | 0 | "e" |
| | Same | 0 | 0 | $p^{R2}$ | $(p^{D2}/(1-\tau))(1-\tau) = p^{D2}$ | "f" |

The first three rows of Table A2.2 show where teachers in an APM school in year 2 were located in year 1. For example, the first row shows that all teachers who came from another APM school must have been movers, since remainers and likers have no reason to switch schools and dislikers are not in APM schools in year 2. The second row shows that teachers who move from non-APM to APM schools are either likers who were randomly assigned to a non-APM schools in year 1 or movers who randomly move from one school to another regardless of schools' APM status. Finally, the third row shows that teachers in an APM school in year 1 who stay there in year 2 are either remainers or likers who were randomly assigned to an APM school in year 1. The fourth, fifth and sixth rows of Table A2.2 show were teachers in non-APM schools in year 2 were located in year 1, and the components of each row follow the same logic as those for the first three rows.

As with Table A2.1, a + b + c = 1 and d + e + f = 1. There are six nonlinear equations here:

$$(1-p^{R2} - p^{L2}/\tau)\mu = a \quad \text{(A2.25)}$$
$$(p^{L2}/\tau)(1-\tau) + (1-p^{R2} - p^{L2}/\tau)(1-\mu) = b \quad \text{(A2.26)}$$
$$p^{L2} + p^{R2} = c \quad \text{(A2.27)}$$
$$(1-p^{R2} - p^{D2}/(1-\tau))\mu + (p^{D2}/(1-\tau))\tau = d \quad \text{(A2.28)}$$
$$(1-p^{R2} - p^{D2}/(1-\tau))(1-\mu) = e \quad \text{(A2,29)}$$
$$p^{R2} + p^{D2} = f \quad \text{(A2.30)}$$

We know the value of τ, which is 3795/6207, but we do not know μ. So we have 6 equations and 4 unknowns: $p^{L2}$, $p^{R2}$, $p^{D2}$ and μ (recall that $p^{M2} = 1 - p^{L2} - p^{R2} - p^{D2}$, so we could add $p^{M2}$, and this equation, to have 7 equations and 5 unknowns).

These 6 equations are not independent. First , adding equations (A2.25) and (A2.26) gives:

$$a + b = (1-p^{R2} - p^{L2}/\tau)\mu + (p^{L2}/\tau)(1-\tau) + (1-p^{R2} - p^{L2}/\tau)(1-\mu) \qquad \text{(A2.31)}$$

$$= \mu - \mu p^{R2} - p^{L2}(\mu/\tau) + (p^{L2}/\tau) - p^{L2} + 1 - p^{R2} - p^{L2}/\tau - \mu + \mu p^{R2} + p^{L2}(\mu/\tau)$$

$$= -p^{L2} + 1 - p^{R2}$$

This implies that:

$$1 - (a + b) = p^{L2} + p^{R2} \qquad \text{(A2.32)}$$

which equals equation (A2.27), since $a + b + c = 1$ (and thus $c = 1 - (a + b)$).

Similarly, adding equations (A2.28) and (A2.29) yields equation (A2.30):

$$d + c = (1 - p^{R2} - p^{D2}/(1-\tau))\mu + (p^{D2}/(1-\tau))\tau + (1 - p^{R2} - p^{D2}/(1-\tau))(1-\mu) \qquad \text{(A2.33)}$$

$$= \mu - \mu p^{R2} - \mu p^{D2}/(1-\tau) + (p^{D2}/(1-\tau))\tau + 1 - p^{R2} - p^{D2}/(1-\tau)) - \mu + \mu p^{R2} + \mu p^{D2}/(1-\tau)$$

$$= (p^{D2}/(1-\tau))\tau + 1 - p^{R2} - p^{D2}/(1-\tau))$$

$$= (p^{D2}/(1-\tau))(\tau-1) + 1 - p^{R2}$$

$$= 1 - p^{R2} - p^{D2}$$

This can be rewritten as:

$$1 - (d + e) = p^{R2} + p^{D2} \qquad \text{(A2.34)}$$

which equals equation (A2.30), since $d + e + f = 1$ (and thus $f = 1 - (d + e)$).

So in fact, we have 4 equations and 4 unknowns. Consider equations (A2.25), (A2.27), (A2.29) and (A2.30). Of these four equations, only two have $\mu$ in them, so they can be used to substitute out $\mu$. From equation (A2.25) we have:

$$\mu = a/(1-p^{R2} - p^{L2}/\tau) \qquad \text{(A2.35)}$$

From equation (A2.29) we have:

$$(1-\mu) = \frac{e}{(1-p^{R2} - p^{D2}/(1-\tau))} \qquad \text{(A2.36)}$$

which implies that:

$$\mu = 1 - \frac{e}{(1-p^{R2} - p^{D2}/(1-\tau))} \qquad \text{(A2.37)}$$

Combining these two equations gives:

$$a/(1-p^{R2} - p^{L2}/\tau) = 1 - \frac{e}{(1-p^{R2} - p^{D2}/(1-\tau))} \qquad \text{(A2.38)}$$

This can be rewritten as:

$$a(1- p^{R2} – p^{D2}/(1-\tau)) = (1- p^{R2} – p^{D2}/(1-\tau))(1-p^{R2} - p^{L2}/\tau) – e(1-p^{R2} - p^{L2}/\tau) \quad \text{(A2.39)}$$

One can substitute out $p^{L2}$ and $p^{D2}$ using equations (A2.27) and (A2.30):

$$a(1- p^{R2} – (f-p^{R2})/(1-\tau)) = (1- p^{R2} – (f-p^{R2})/(1-\tau))(1-p^{R2} – (c-p^{R2})/\tau) \quad \text{(A2.40)}$$

$$– e(1-p^{R2} – (c-p^{R2})/\tau)$$

This gives a quadratic equation in $p^{R2}$:

$$a – af/(1-\tau) + ap^{R2}/(1-\tau) – ap^{R2} = (1 – p^{R2} – c/\tau + p^{R2}/\tau) – p^{R2} + (p^{R2})^2 + cp^{R2}/\tau \quad \text{(A2.41)}$$

$$– (p^{R2})^2/\tau – f/(1-\tau) + fp^{R2}/(1-\tau) + fc/[\tau(1-\tau)] – p^{R2}f/[\tau(1-\tau)] + p^{R2}/(1-\tau)$$

$$– (p^{R2})^2/(1-\tau) – cp^{R2}/[\tau(1-\tau)] + (p^{R2})^2/[\tau(1-\tau)] – e + ep^{R2} + ec/\tau – ep^{R2}/\tau$$

Putting like terms together gives:

$$(p^{R2})^2\{1/\tau – 1 + 1/(1-\tau) – 1/[\tau(1-\tau)]\} \quad \text{(A2.42)}$$

$$+ p^{R2}\{a/(1-\tau) – a + 2 – 1/\tau – c/\tau – f/(1-\tau) + f/[\tau(1-\tau)] – 1/(1-\tau) + c/[\tau(1-\tau)] – e + e/\tau\}$$

$$+ \{a – af/(1-\tau) – 1 + c/\tau + f/(1-\tau) – fc/[\tau(1-\tau)] + e – ec/\tau\} = 0$$

Simplifying terms gives:

$$-(p^{R2})^2 + p^{R2}\{2 – 1/[\tau(1-\tau)] + a\tau/(1-\tau) + c/(1-\tau) + e(1-\tau)/\tau + f/\tau\} \quad \text{(A2.43)}$$

$$+ \{(a + c/\tau)[1 – f/(1-\tau)] – 1 + f/(1-\tau) + e(1 + 1/\tau)\} = 0$$

Simplifying further gives:

$$-(p^{R2})^2 + p^{R2}\{2 – 1/[\tau(1-\tau)] + (a\tau + c)/(1-\tau) + e(1-\tau)/\tau + f/\tau\} \quad \text{(A2.44)}$$

$$+ \{(1 – c/\tau)[f/(1-\tau) + e - 1] + a[1 – f/(1-\tau)]\} = 0$$

Recalling the value of $\tau$ and calculating a, b, c, d, e and f using the numbers from Table 4 in the main text yields:[23]

$$\tau = 3795/6207 = 0.611 \quad \text{(A2.45)}$$

$$a = 33/272 = 0.121 \quad \text{(A2.46)}$$
$$b = 60/272 = 0.221 \quad \text{(A2.47)}$$
$$c = 179/272 = 0.658 \quad \text{(A2.48)}$$
$$d = 34/309 = 0.110 \quad \text{(A2.49)}$$
$$e = 75/309 = 0.243 \quad \text{(A2.50)}$$

[23] For the total number of teachers in Table 4 we exclude “others” since their school of origin is unknown.

$$f = 200/309 = 0.674 \qquad \text{(A2.51)}$$

Applying (A2.45) - (A2.51) to the coefficient on $p^{R2}$ in equation (A2.44) gives:

$$2 - 1/[\tau(1-\tau)] + (a\tau + c)/(1-\tau) + e(1-\tau)/\tau + f/\tau \qquad \text{(A2.52)}$$

$$= 2 - 1/(0.611*0.389) + (0.121*0.611 + 0.658)/0.389 + 0.243*(0.389/0.611) + 0.674/0.611$$

$$= 0.932$$

Applying the same equations to the "constant term" in equation (A2.44) yields:

$$(1 - c/\tau)[f/(1-\tau) + e - 1] + a[1 - f/(1-\tau)] \qquad \text{(A2.53)}$$

$$= (1 - (0.658/0.611))*(0.674/0.389 + 0.243 - 1) + 0.121*(1 - 0.674/0.389)$$

$$= -0.164$$

Inserting (A2.52) and (A2.53) into equation (A2.44) and applying the quadratic formula gives:

$$p^{R2} = \{-0.932 \pm [0.932^2 + 4*(-0.164)]^{1/2}\}/(-2) \qquad \text{(A2.54)}$$

$$= \{-0.932 \pm [0.213]^{1/2}\}/(-2)$$

$$= \{-0.932 \pm 0.461\}/(-2)$$

$$= 0.236 \text{ and } 0.697$$

First consider what happens if $p^{R2} = 0.236$:

By equation (A2.27) we have $p^{L2} + p^{R2} = c$, which implies that:

$$p^{L2} = 0.658 - 0.236 = 0.422 \qquad \text{(A2.55)}$$

By equation (A2.30) we have $p^{R2} + p^{D2} = f$, which implies that:

$$p^{D2} = 0.674 - 0.236 = 0.438 \qquad \text{(A2.56)}$$

Movers must be the remaining category, which is:

$$1 - (0.236 + 0.422 + 0.438) = -0.096 \qquad \text{(A2.57)}$$

Finally, $\mu$ (fraction of movers coming from treated schools) is:

$$a/(1-p^{R2} - p^{L2}/\tau) \qquad \text{(A2.58)}$$

$$= 0.121/(1 - 0.236 - 0.422/0.611)$$

$$= 1.651$$

These results for $p^{M2}$ and $\mu$ are nonsensical, so consider $p^{R2} = 0.697$. By equation (A2.27) we have $p^{L2} + p^{R2} = c$, which implies that:

$$p^{L2} = 0.658 - 0.697 = -0.039 \qquad \text{(A2.59)}$$

By equation (A2.30) we have $p^{R2} + p^{D2} = f$, which implies that:

$$p^{D2} = 0.674 - 0697 = -0.023 \qquad \text{(A2.60)}$$

Movers must be the remaining category, which is:

$$1 - (0.697 - 0.039 - 0.023) = 0.365 \qquad \text{(A2.61)}$$

Finally, $\mu$ (fraction of movers coming from treated schools) is:

$$a/(1-p^{R2} - p^{L2}/\tau) \qquad \text{(A2.62)}$$

$$= 0.121/(1 - 0.697 - (-0.039/0.611)$$

$$= 0.330.$$

# Online Appendix 3. Additional Tables and Figures

**Table A3.1**
**Selective Attrition Test: Regression of Treatment Status on Pre-treatment Characteristics**

| | (1) All teachers | (2) Observed teachers at the end of 2017 | (3) Non-observed teachers at the end of 2017 |
|---|---|---|---|
| Age | 0.001 | 0.003 | -0.000 |
| | (0.003) | (0.003) | (0.005) |
| Male | -0.001 | 0.022 | -0.107 |
| | (0.038) | (0.051) | (0.090) |
| *Education* | | | |
| Higher education | -0.059 | 0.069 | -0.302 |
| | (0.183) | (0.095) | (0.353) |
| Postgraduate | -0.102 | 0.027 | --- |
| | (0.194) | (0.127) | |
| *Teacher career* | | | |
| Contract | 0.105 | 0.059 | 0.158 |
| | (0.066) | (0.082) | (0.130) |
| 2nd scale | -0.017 | 0.024 | -0.066 |
| | (0.059) | (0.075) | (0.117) |
| 3rd scale | 0.110 | 0.088 | 0.180 |
| | (0.072) | (0.092) | (0.141) |
| 4th scale | -0.015 | 0.048 | -0.213 |
| | (0.091) | (0.123) | (0.258) |
| 5th scale | -0.037 | 0.040 | --- |
| | (0.159) | (0.197) | |
| Joint Significance Test - pvalue | 0.507 | 0.942 | 0.136 |
| N | 646 | 444 | 202 |
| R-squared | 0.212 | 0.244 | 0.440 |

Note: All regressions include UGEL fixed effects. Standard errors clustered at the school level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

**Table A3.2**
**Selective Attrition Test: Regression of Observed Indicator on Treatment Status, Pre-treatment Characteristics, and Interactions of Both Variables**

| | Observed in Year 2 (Yes=1) | |
|---|---|---|
| Treatment | -0.706* | (0.408) |
| Treatment x Age | 0.007 | (0.005) |
| Treatment x Male | -0.040 | (0.082) |
| Treatment x Higher Education | 0.403 | (0.338) |
| Treatment x Postgraduate | 0.271 | (0.350) |
| Treatment x Contract | -0.029 | (0.115) |
| Treatment x 2nd scale | 0.075 | (0.100) |
| Treatment x 3rd scale | -0.054 | (0.121) |
| Treatment x 4th scale | -0.007 | (0.189) |
| Treatment x 5th scale | 0.318 | (0.247) |
| Age | -0.003 | (0.003) |
| Male | 0.008 | (0.058) |
| Higher education | 0.049 | (0.295) |
| Postgraduate | 0.344 | (0.303) |
| Contract Teacher | -0.153* | (0.086) |
| 2nd scale | -0.068 | (0.070) |
| 3rd scale | -0.023 | (0.082) |
| 4th scale | -0.077 | (0.154) |
| 5th scale | 0.016 | (0.121) |
| Joint Significance Test – pvalue (treatment and interactions between treatment and pre-treatment characteristics) | 0.373 | |
| N | 646 | |
| R-squared | 0.255 | |

Note: The regression includes UGEL fixed effects. Standard errors clustered at the school level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

**Table A3.3**
**Heterogeneous Treatment Effects in Sample 1**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Treatment | 0.314*** | 0.213 | 0.314*** | 0.273* | 0.236 |
| | (0.102) | (0.240) | (0.117) | (0.159) | (0.147) |
| Experience | 0.000 | -0.002 | 0.000 | 0.000 | 0.000 |
| | (0.009) | (0.011) | (0.009) | (0.009) | (0.009) |
| Contract Teacher | 0.152 | 0.153 | 0.154 | 0.154 | 0.140 |
| | (0.162) | (0.163) | (0.226) | (0.163) | (0.162) |
| Magisterial Level | 0.114** | 0.115** | 0.114** | 0.102* | 0.114** |
| | (0.046) | (0.046) | (0.046) | (0.060) | (0.046) |
| Sex (Men=1) | -0.313*** | -0.315*** | -0.313*** | -0.313*** | -0.396*** |
| | (0.099) | (0.099) | (0.099) | (0.099) | (0.147) |
| Age | -0.029*** | -0.029*** | -0.029*** | -0.029*** | -0.029*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| Treatment #Experience | | 0.005 | | | |
| | | (0.011) | | | |
| Treatment #Contract | | | -0.004 | | |
| | | | (0.247) | | |
| Treatment #M. Level | | | | 0.025 | |
| | | | | (0.081) | |
| Treatment #Sex | | | | | 0.170 |
| | | | | | (0.188) |
| $R^2$ | 0.37 | 0.37 | 0.37 | 0.37 | 0.37 |
| $N$ | 455 | 455 | 455 | 455 | 455 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Note: All regressions include UGEL fixed effects. Standard errors clustered at the school level are reported in parenthesis.

**Table A3.4.**
**Intention-to-Treat Estimates on Mathematics Test Scores**

| VARIABLES | 2nd grade (2016) | | 4th grade (2016) | | 2nd and 4th grade (2016) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment (Yes=1) | 0.259* | 0.188 | 0.169 | 0.595** | 0.181* | 0.338** |
| | (0.135) | (0.206) | (0.157) | (0.281) | (0.102) | (0.151) |
| Treatment×Number of Teachers | | 0.0604 | | -0.152* | | -0.0309 |
| | | (0.0674) | | (0.0791) | | (0.0488) |
| Number of teachers | | 0.0936** | | -0.0021 | | 0.0648** |
| | | (0.0398) | | (0.0403) | | (0.0302) |
| Constant | 0.394*** | 0.0693 | -0.124 | -0.117 | -0.479* | -0.719** |
| | (0.0786) | (0.178) | (0.469) | (0.500) | (0.273) | (0.317) |
| Observations | 1,339 | 1,269 | 1,184 | 1,125 | 2,523 | 2,394 |
| Number of clusters | 151 | 138 | 140 | 129 | 181 | 161 |
| R-squared | 0.269 | 0.291 | 0.305 | 0.315 | 0.446 | 0.461 |
| UGEL FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade FE | No | No | No | No | Yes | Yes |

Note: Estimates are in standard deviations (s.d.). Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table A3.5.**
**Intention-to-Treat Estimates on Reading Test Scores**

| VARIABLES | 2nd grade (2016) | | 4th grade (2016) | | 2nd and 4th grade (2016) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment (Yes=1) | 0.240* | 0.0944 | 0.111 | 0.573** | 0.141 | 0.284** |
| | (0.123) | (0.207) | (0.153) | (0.255) | (0.0903) | (0.138) |
| Treatment×Number of teachers | | 0.0901 | | -0.155* | | -0.0255 |
| | | (0.0674) | | (0.0795) | | (0.0467) |
| Number of teachers | | 0.0780* | | 0.0284 | | 0.0513* |
| | | (0.0398) | | (0.0390) | | (0.0268) |
| Constant | 0.476 | 0.206 | -0.347 | -0.451 | -0.661 | -0.852* |
| | (0.376) | (0.428) | (0.591) | (0.628) | (0.469) | (0.498) |
| Observations | 1,339 | 1,269 | 1,185 | 1,126 | 2,524 | 2,395 |
| Number of clusters | 151 | 138 | 140 | 129 | 181 | 161 |
| R-squared | 0.286 | 0.312 | 0.255 | 0.256 | 0.503 | 0.513 |
| UGEL FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade FE | No | No | No | No | Yes | Yes |

Note: Estimates are in standard deviations (s.d.). Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1